



**HAL**  
open science

## A Wide Lens

Kristine Lund, Gerald P. Niccolai, Cindy Hmelo-Silver, Gweonho Jeong,  
Michael Baker

► **To cite this version:**

Kristine Lund, Gerald P. Niccolai, Cindy Hmelo-Silver, Gweonho Jeong, Michael Baker. A Wide Lens. 2019. hal-03990455

**HAL Id: hal-03990455**

**<https://telecom-paris.hal.science/hal-03990455v1>**

Submitted on 15 Feb 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Public Domain



13<sup>TH</sup>  
LYON

CSCL 2019



**A Wide Lens:**

**Combining Embodied, Enactive, Extended, and  
Embedded Learning in Collaborative Settings:**

**13<sup>th</sup> International Conference on Computer Supported Collaborative Learning**

**Conference Proceedings Volume 1**

**Edited by**

**Kristine Lund, Gerald Niccolai, Elise Lavoué, Cindy Hmelo-Silver  
Gahgene Gweon, Michael Baker**

**17 - 21 June 2019**

**École Normale Supérieure de Lyon, France**







**A Wide Lens: Combining Embodied, Enactive,  
Extended, and Embedded Learning in  
Collaborative Settings**

13<sup>th</sup> International Conference on Computer  
Supported Collaborative Learning (CSCL) 2019

Volume 1

13<sup>th</sup> International Conference on  
Computer Supported Collaborative Learning

June 17-21, 2019, Lyon, France  
École Normale Supérieure de Lyon, France

Editors:

Kristine Lund, Gerald Niccolai, Elise Lavoué, Cindy Hmelo-Silver  
Gahgene Gweon, and Michael Baker

*A Wide Lens: Combining Embodied, Enactive, Extended, and Embedded Learning in Collaborative Settings, 13th International Conference on Computer Supported Collaborative Learning (CSCL) 2019*

June 17-21, 2019, Lyon, France

École Normale Supérieure de Lyon, France.

© 2019 International Society of the Learning Sciences, Inc. [ISLS]. Rights reserved.



[www.isls.org](http://www.isls.org)

ISBN: 978-1-7324672-3-1 (Volume 1, PDF Version)

ISSN: 1573-4552

Cite as: Lund, K., Nicolai, G., Lavoué, E., Hmelo-Silver, C., Gweon, G., and Baker, M. (Eds.). (2019). *A Wide Lens: Combining Embodied, Enactive, Extended, and Embedded Learning in Collaborative Settings, 13th International Conference on Computer Supported Collaborative Learning (CSCL) 2019, Volume 1*. Lyon, France: International Society of the Learning Sciences.

All rights reserved. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear the copyright notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior written permission of the International Society of the Learning Sciences. The International Society of the Learning Sciences is not responsible for the use which might be made of the information contained in this book.

# Conference Organizers and Committees

## Conference Chair and Co-Chairs

Kristine Lund, *CNRS, ENS Lyon, University of Lyon 2, France*  
Gerald P. Niccolai, *CNRS, ENS Lyon, University of Lyon 2, France*  
Elise Lavoué, *Jean Moulin Lyon 3 University, France*

## Program Committee Co-Chairs

Cindy Hmelo-Silver, *Indiana University, USA*  
Gahgene Gweon, *Seoul National University, South Korea*  
Michael Baker, *CNRS - Telecom Paris-Tech, France*

## Social Media Communications Co-Chairs

Bodong Chen, *University of Minnesota, USA*  
Dimitra Tsovaltzi, *University of Saarbrücken, Germany*

## Doctoral Consortium Co-Chairs

Heisawn Jeong, *Hallym University, South Korea*  
Yotam Hod, *University of Haifa, Israel*  
Kate Bielaczyc, *Clark University, Massachusetts, USA*

## Early Career Co-Chairs

Julia Eberle, *Bochum University, Germany*  
Victor Lee, *Utah State University, USA*  
Nancy Law, *Hong Kong University*

## Mid-Career Chair

Stephanie Teasley, *University of Michigan, USA*  
Marcia Linn, *University of California at Berkeley, USA*  
Janet Kolodner, *Georgia Institute of Technology, USA*  
Susan Goldman, *University of Illinois at Chicago, USA*

## Interactive Tools and Demos Co-Chairs

Karim Sehaba, *University of Lyon 2, France*  
Yannis Dimitriadis, *University of Valladolid, Spain*  
Pierre-Antoine Champin, *University of Lyon 1, France*

## Workshops and Tutorials Co-Chairs

Freydis Vogel, *University of Nottingham, UK*  
Sherice Clarke, *University of California at San Diego, USA*  
Andrew Joyce Gibbons, *Durham University, UK*  
Lenka Schnaubert, *University of Duisburg-Essen*

## Current Significant Discussions and the Future of ijCSCL

Sten Ludvigsen, *University of Oslo, Norway*  
Alyssa Wise, *New York University, NY, USA*  
Baruch B. Schwarz, *Hebrew University of Jerusalem, Israel*

## What 4E Cognition Can Tell Us About the Successes and Failures of CSCL Implementation

Gaëlle Molinari, *University of Geneva, Switzerland*  
Stian Håklev, *École Polytechnique Fédérale de Lausanne, Switzerland*  
Claire Polo, *University of Lyon 2, France*

## **Advisory Committee**

Marcela Borge, *Pennsylvania State University, USA*

Oskar Lindwall, *University of Gothenburg, Sweden*

Paul Kirschner, *Open Universiteit Nederland, Heerlen, the Netherlands*

Sten Ludvigsen, *University of Oslo, Norway*

Sadhana Puntambekar, *University of Wisconsin at Madison, USA*



## Senior Reviewers

Jan van Aalst  
Richard Alterman  
Michael Baker  
François-Xavier Bernard  
Marcela Borge  
Susan Bridges  
Éric Bruillard  
Murat Cakir  
Nur Akkus Cakir  
Carol Chan  
Wenli Chen  
Britte Cheng  
Sherice Clarke  
Ulrike Cress  
Cynthia D'Angelo  
Crina Damsa  
Joshua Danish  
Erica de Vries  
Bram De Wever  
Yannis Dimitriadis  
Sébastien George  
Gahgene Gweon  
Kai Hakkarainen  
Päivi Häkkinen  
Cindy Hmelo-Silver  
Christopher Hoadley  
Yotam Hod  
H. Ulrich Hoppe  
Heisawn Jeong  
Timothy Koschmann  
Kristiina Kumpulainen

Nancy Law  
Victor Lee  
Kyu Yon Lim  
Oskar Lindwall  
Chee-Kit Looi  
Sten Ludvigsen  
Jean-Charles Marty  
Toshio Mochizuki  
Anders Mørch  
Hanni Muukkonen  
Omid Noroozi  
Jun Oshima  
Kylie Peppler  
Palmyre Pierroux  
Joseph Polman  
Sadhana Puntambekar  
Christophe Reffay  
Peter Reimann  
Carolyn Rosé  
Nikol Rummel  
Baruch Schwarz  
Hajime Shirouzu  
Nancy Songer  
Gerry Stahl  
Seng Chee Tan  
Pierre Tchounikine  
Freydis Vogel  
Armin Weinberger  
Alyssa Wise  
Jason Yip  
Susan Yoon

## Reviewers

Pei-Yi Lin  
Giovanna Scalone  
Dor Abrahamson  
Alejandro Andrade  
Deborah Fields  
Yacine Lafifi  
Richard Medina  
Amando Singun  
Christopher Thorn  
Michael Tscholl  
Lucas Bietti  
Elizabeth Charles  
Pryce Davis  
Sanna Jarvela  
Yael Kali  
Niki Lambropoulos  
Heather Leary  
Alwyn Vwen Yen Lee  
Jasmine Ma  
Jennifer Olsen  
Johanna Poysa-Tarhonen  
Amanda Siebert-Evenstone  
Patrick Sunnen  
Joanna Weidler-Lewis  
Mireille Betrancourt  
Carrie Demmans Epp  
Gijsbert Erkens  
Danielle Espino  
Susan Goldman  
Erica Halverson  
Maria Antonietta Impedovo  
Aaron Kessler  
Brian Krisler  
Feng Lin  
Stefania Manca  
Frederick Peck  
Christoph Richter  
Giuseppe Ritella  
Toni Rogat  
Asmalina Saleh  
David Shaenfield  
Mike Tissenbaum  
Hakan Tuzun  
Philip Vahey  
Tugce Aldemir  
Huiying Cai  
Clark Chinn  
Gi Woong Choi  
Andras Csanadi  
Jan Arild Dolonen  
Jonan Donaldson  
Matthew Duvall  
Julia Eberle  
Howard Everson  
Amy Fox  
Engida Gebre  
Insook Han  
So Hyo-Jeong  
Allan Jeong  
Priscilla Jimenez Pazmino  
Ilias Karasavvidis  
Anna Keune  
Andrea Kienle  
Yanghee Kim  
Samuel Kwon  
Eleni Kyza  
Minna Lakkala  
Luettamae Lawrence  
Chien-Sing Lee  
Teemu Leinonen  
Leanne Ma  
Lina Markauskaite  
Magdalene Moy  
Amanda Ochsner  
Hiroki Oura  
Fan Ouyang  
Teresa Pargman  
Sambit Praharaaj  
Daniel Rees Lewis  
Jessica Roberts  
Mamta Shah  
Karsten Stegmann  
Florence Sullivan  
Hamideh Talaftian  
Dimitra Tsovaltzi  
Charles Vanover  
Hanna Vuojarvi  
Xu Wang  
Chaolun Xia  
Gaoxia Zhu  
Christa Asterhan  
Afaf Baabdullah  
Juergen Buder  
Gaowei Chen  
Charlie Cox  
Mutlu Cukurova  
Kiran Eranki  
Nobuko Fujita  
Andrea Gomoll  
Arnseth Hans Christian  
Jiangang Hao  
Kian Sam Hong  
Anesa Hosein  
Frank Jong  
Yong Ju Jung  
Chanmin Kim  
Paul Kirschner  
Elizabeth Koh  
Susan Land  
Andreas Lingnau  
Alejandra Martínez-Monés  
Mitchell Nathan  
Juliana Nazare  
Ha Nguyen  
Nicolae Nistor  
Martina Rau  
Traian Rebedea  
Mimi Recker  
Norma Reichelt  
Kristin Searle  
Pirita Seitamaa-Hakkarainen  
Paul Seitlinger  
Gerard Sensevy  
Ji Shen  
Jim Slotta  
Sebastian Strauß  
Jan-Willem Strijbos  
Zachari Swiecki  
Kate Thompson  
Judith Uchidiuno  
Xinghua Wang  
Joshua Weese  
Marcelo Worsley  
Jimmy Zambrano R.  
Robert Ashley  
Tutaleni I. Asino  
Na'ama Av-Shalom  
Katerine Bielaczyc  
Daniel Bodemer  
Alexander Brooks  
Cynthia Carter Ching  
Min Chen  
John Cherniavsky  
Etan Cohen  
David Deliema  
Stavros Demetriadis  
Rachel Dickler  
Bernhard Ertl  
Abigail Evans  
Georgios Fesakis  
Paul Flynn  
Dilrukshi Gamage  
Maria Teresa Gastardo  
Aristotelis Gkiolmas  
Frode Guribye  
Thomas Hillman  
Ty Hollett  
Joey Huang  
Roland Hubscher  
Dongsik Kim  
Simon Knight  
Ingo Kollar

Elise Lavoué  
Xiao Li  
Roberto Martinez-Maldonado  
Eileen Mcgivney  
Gaelle Molinari  
Isaac Nichols-Paez  
Annemarie Palincsar  
Pantelis M. Papadopoulos  
Isidoros Perikos  
Ashlyn Pierson  
Ingvill Rasmussen  
Ornit Sagy  
Lay Hoon Seah  
D.W. Shaffer  
Tanmay Sinha  
Ole Smørdal  
Trausan-Matu Stefan  
Anouschka Van Leeuwen  
Michiel Voet  
Patrick Wang  
Yun Wen  
Longkai Wu  
Shulong Yan  
Dejin Zhao  
Katelyn Alderfer  
Maria Altebarmakian  
Golnaz Arastoopour Irgens  
Francine Athias  
Daniel Belenky  
Yifat Ben-David Kolikant  
Dani Ben-Zvi  
Yoav Bergner  
Dario Bertero  
Spencer Carlson  
Ming Chiu  
Irene-Angelica Chounta  
Tamara Clegg  
Mihai Dascalu  
Sharon Derry  
Michal Dvir  
Erkan Er  
Xueqi Feng  
Lin Feng  
Barry Fishman  
Erin Furtak  
Michael Glass  
Judith Green  
Christian Greiffenhagen  
Dan Hickey  
Megan Humburg  
Shinya Iikubo  
Michael Jacobson  
Kaiju Kangas  
Anastasios Karakostas  
Mi Song Kim  
Beaumie Kim  
Soo Hyeon Kim

Allison Kolling  
Tiina Korhonen  
Ingeborg Krange  
Therese Laferriere  
Beatrice Ligorio  
Rasheda Likely  
Patrik Lilja  
Robin Lin  
Gabriela Locoh-A  
Katharina Loibl  
Guanzhong Ma  
Anthony Matranga  
Tom Moher  
Ritsuko Oshima  
Seung Won Park  
Caroline Pitt  
Josh Radinsky  
Annelies Raes  
Gabriela Richard  
Ricarose Roque  
Rosemary Russ  
Stéphane Safin  
Lara Schmitt  
Bertrand Schneider  
Kshitij Sharma  
Saadeddine Shehab  
Patrick Shih  
Daniel Spikol  
Rolf Steier  
Jingjing Sun  
Chew Lee Teo  
Jakob Tholander  
Yuyao Tong  
Dhvani Toprani  
Yu Xia  
Katerina Zourou  
Rotem Abdu  
Cengiz Acarturk  
Rodica Ailincai  
Martha Alibali  
Jerry Andriessen  
Golnaz Arastoopour Irgens  
Georges-Louis Baron  
Matthew Berland  
Elizabeth Bonsignore  
Ivica Botički  
Linda Bradley  
Corey Brady  
Julien Broisin  
Fabio Campos  
Lucila Carvalho  
Tieh-Huai Chang  
Monaliza Chian  
Matthieu Cisel  
Ashley Colley  
Sugat Dabholkar  
Jacob Davidsen

Bria Davis  
Jort De Vreeze  
Chris Dede  
Anne Deiglmayr  
Betsy Disalvo  
Sidney Dmello  
Catherine Dornfeld  
Tissenbaum  
Toya Frank  
Luke Fryer  
Chris Georgen  
Dana Gnesdilow  
Shuchi Grover  
Joerg Haake  
Stian Håklev  
Rajja Hämäläinen  
Tellervo Härkki  
Christian Hartmann  
Katie Headrick Taylor  
Sven Heimbuch  
Lauri Hietajärvi  
Jaakko Hilppö  
Peter Holtz  
Sameer Honwad  
Mike Horn  
Helen Hu  
Diane Hui  
Jonas Ivarsson  
Lara Jasien  
Kuan-Yu Jin  
Emily Johnson  
Mina Johnson  
A. Susan Jurow  
Kalle Juuti  
Jennifer Kahn  
Jina Kang  
Shiri Kashi  
Fengfeng Ke  
Susan Kelly  
Takeshi Kitazawa  
Eric Klopfer  
Vassilis Komis  
Siu Cheung Kong  
Deanna Kuhn  
Huei-Yi Lai  
Jari Lavonen  
Shupin Li  
Yongyan Li  
Kenneth Lim  
Robb Lindgren  
Breanne Litts  
Lei Liu  
Kristine Lu  
Gustav Lymer  
Leilah Lyons  
Emmanuel Manalo  
Andrew Manches

Miika Marttunen  
Hiroyuki Masukawa  
Camillia Matuk  
Steven Mcgee  
Julien Mercier  
Mikko Meriläinen  
Anahid Modrek  
Lorenza Mondada  
Chrystalla Mouza  
Kae Nakaya  
Markus Nivala  
Xavier Ochoa  
Tuire Palonen  
Anthony Pellicone  
Nathan Phillips  
Claire Polo  
Kristina Popova  
Luis P. Prieto  
Michael Prilla

Rebecca Quintana  
Chris Quintana  
Liat Rahmian  
Daniel Reinholz  
Symeon Retalis  
Katie Rich  
Margarida Romero  
Tijs Rotsaert  
Jean Ryoo  
Moegi Saito  
Sandra Schulz  
R. Mishael Sedas  
Priya Sharma  
Brett Shelton  
Anna Shvarts  
Audra Skukauskaite  
Erica Snow  
Kiley Sobel  
Firat Soylu

Jin Splichal  
Jin Splichal  
Naomi Thompson  
Xintian Tu  
Michael Twidale  
Ralph Vacca  
Abha Vaishampayan  
Himanshu Verma  
Devan Walton  
Maggie Wang  
Tsering Wangyal  
Michael Weinstock  
Korah Wiley  
Iwan Wopereis  
Yuqin Yang  
Junxiu Yu  
Lily Zeng  
Si Zhang  
Heather Toomey Zimmerman

# Preface

On behalf of the whole CSCL 2019 organizing team, we are delighted to welcome you to Lyon! Our theme is

**A Wide Lens: Combining Embodied, Enactive, Extended, and Embedded Learning in Collaborative Settings.**

Promoting productive collaborative interaction in varied contexts requires studying the interdependencies of the complex ecosystems in which collaborative learning takes place (e.g. school, museums, work, play). In Cognitive Science, research under the banner of 4E cognition favors newer paradigms that take into account the role of the body (embodied), the interactions between an organism and its environment (enactive), and the elements and aspects in the environment itself (extended and embedded). At this year's CSCL, we propose 4E *learning* as our theme. Submissions that present results on collaborative learning regarding some combination of embodied, enactive, extended, and embedded learning are welcome. Such a focus translates to studies of various interdependencies in the learning process: social, emotional, cultural, linguistic, cognitive, and technological. Finally, treating 4E learning as inherently collaborative means that as a CSCL community, we need to understand this phenomenon in settings both with and without technology. It is important that as CSCL researchers, we build on work that while not computer-supported has implications for design and research in computer-supported collaborative learning settings. In considering collaborative learning as the interplay of factors in a complex system, we aim to create novel interdisciplinary integrations and thereby extend and reinforce the CSCL Learning Sciences community with new ideas.

We hope you enjoy the program as well as visiting our beautiful city!

Kris Lund, conference chair CSCL 2019 (on behalf of the whole team)



## Acknowledgments

CSCL 2019 organizers would like to thank all of our sponsors for their support: ENS de Lyon, Université de Lyon, IDEX Lyon, CNRS, Université Jean-Moulin Lyon 3, la Région Auvergne-Rhône-Alpes, Université Lumière Lyon 2, l'INSA, la ville de Lyon, le LabEx ASLAN, NSF, and ISLS, as well as the research laboratories ICAR and LIRIS. We have appreciated working with Insight-Outside, Professional Conference Organizer, in charge of certain aspects of our event's logistics. Thank-you also to the volunteers and to Santiago Guillen for organizing them. The conference logo was designed by François Gentil and the cover design was a collaborative effort between Insight-Outside, Gerald Niccolai, and Allison Hall. Many thanks to Allison for putting together the proceedings and dealing with last minute changes to the program.

# Table of Contents

## Volume 1

### Keynotes

|   |   |
|---|---|
| Integrating Theorizing on Embodied, Enactive, Extended, and Embedded Cognition to Augment CSCL Research ..... | 3 |
| <i>Stephen Fiore</i>  |   |
| Negotiating Knowledge, Expertise, Connoisseurship and Taste in Social Interaction .....                       | 4 |
| <i>Lorenza Mondada</i>  |   |
| Designing Opie Robots as Learning Companions: Insights, Interactions and Interdependencies .....              | 5 |
| <i>Janet Wiles</i>  |   |

### Full Papers

|   |     |
|---|-----|
| Analysis of Touchscreen Interactive Gestures During Embodied Cognition in Collaborative Tabletop Science Learning Experiences .....   | 9   |
| <i>Nikita Soni, Alice Darrow, Annie Luc, Schuyler Gleaves, Carrie Schuman, Hannah Neff, Peter Chang, Brittani Kirkland, Jeremy Alexandre, Amanda Morales, Kathryn A. Stofer, Lisa Anthony</i> |     |
| “I Cannot Explain Why I Like This Shape Better Than That Shape”: Intercorporeality in Collaborative Learning .....  | 17  |
| <i>Jacob Davidsen, Thomas Ryberg</i>  |     |
| Effects of the Need for Cognitive Closure and Guidance on Contribution Quality in Wiki-Based Learning .....   | 25  |
| <i>Sven Heimbuch, Daniel Bodemer</i>  |     |
| Expansively Framing Social Annotations for Generative Collaborative Learning in Online Courses .....  | 33  |
| <i>Christopher Andrews, Grant Chartrand, Daniel T. Hickey</i>   |     |
| Unpacking Collaborative Learning Processes During Hands-on Activities Using Mobile Eye-Trackers .....   | 41  |
| <i>Bertrand Schneider</i>   |     |
| The Lifespan and Impact of Students’ Ideas Shared During Classroom Science Inquiry .....  | 49  |
| <i>Camillia Matuk, Wanjing (Anyu) Ma, Garima Sharma, Marcia Linn</i>  |     |
| Collective Knowledge Advancement through Shared Epistemic Agency: Socio-Semantic Network Analyses .....   | 57  |
| <i>Jun Oshima, Ritsuko Oshima, Ayano Ohsaki, Jin Splichal</i>   |     |
| Families and Media Multi-tasking: Reorganizing Collaborative Learning at Home .....   | 65  |
| <i>Deborah Silvis</i>   |     |
| Designing Representations in Deeply Disciplinary Educational Games .....  | 73  |
| <i>Ravit Golan Duncan, Frieda Reichsman, Kiley McElroy-Brown, Veronica Cavera</i>   |     |
| The Knowledge Building Network Pilot Project: An Exploration of Emergent Designs to Enhance Collective Teacher Efficacy .....   | 81  |
| <i>Leanne Ma, Monica Resendes, Marlene Scardamalia, Karen Dobbie</i>  |     |
| Comparing the Effectiveness of Supports for Collaborative Dialogic Sense-Making with Agent-Based Models .....   | 88  |
| <i>Ilana Dubovi, Victor Lee</i>   |     |
| Using Voice Assistant Skills in Family Life .....   | 96  |
| <i>Diana Beirl, Nicola Yuill, Yvonne Rogers</i>   |     |
| Measuring the Computational in Computational Participation: Debugging Interactive Stories in Middle School Computer Science .....   | 104 |

|  |     |
|--|-----|
| <i>Chris Proctor</i>   |     |
| Does Order Matter? Investigating Sequential and Cotemporal Models of Collaboration .....                                       | 112 |
| <i>Zachari Swiecki, Zheming Lian, Andrew Ruis, D.W. Shaffer</i>  |     |
| Teacher Monitoring Routines: Understanding Pedagogical Judgments During Students’<br>Collaborative Learning .....              | 120 |
| <i>Nadav Ehrenfeld, Ilana Horn</i>   |     |
| Impacts of Augmented Reality on Collaborative Physics Learning, Leadership, and<br>Knowledge Imbalance .....                   | 128 |
| <i>Iulian Radu, Bertrand Schneider</i>   |     |
| Finding Common Ground: A Method for Measuring Recent Temporal Context in Analyses of Complex,<br>Collaborative Thinking .....  | 136 |
| <i>Andrew Ruis, Amanda Siebert-Evenstone, Rebecca Pozen, Brendan Eagan, D.W. Shaffer</i>                                       |     |
| Idea Thread Mapper: Designs for Sustaining Student-Driven Knowledge Building Across Classrooms .....                           | 144 |
| <i>Jianwei Zhang, Mei-Hwa Chen</i>   |     |
| Location, Location, Location: The Effects of Place in Place-Based Simulations.....   | 152 |
| <i>Amanda Siebert-Evenstone, D.W. Shaffer</i>  |     |
| The Role of Funds of Knowledge in Online Search and Brokering .....  | 160 |
| <i>Wendy Roldan, Paola Vanegas, Laura Pina, Carmen Gonzalez, Jason Yip</i>   |     |
| Collaboration on a Massive Scale – Conceptual Implications of the Crowd.....   | 168 |
| <i>Norma Reichelt, Christine Bussian, Christoph Richter, Heidrun Allert, Lars-Arne Raffel</i>                                  |     |
| Adaptive Support for Collaboration on Tabletop Computers .....   | 176 |
| <i>Abigail Evans, Katie Davis, Jacob Wobbrock</i>  |     |
| Integrative Visualization: Exploring Data Collected in Collaborative Learning Contexts.....                                    | 184 |
| <i>Ben Shapiro</i>   |     |
| An Empirical Study of Educational Robotics as Tools for Group Metacognition and<br>Collaborative Knowledge Construction.....   | 192 |
| <i>Chrysanthos Socratous, Andri Ioannou</i>  |     |
| Disciplinary Task Models for Designing Classroom Orchestration: The Case of Data Visualization for<br>Historical Inquiry.....  | 200 |
| <i>Josh Radinsky, Iris Tabak, Monika Moore</i>   |     |
| Data Wrangling Practices and Process in Modeling Family Migration Narratives with Big Data<br>Visualization Technologies ..... | 208 |
| <i>Shiyang Jiang, Jennifer Kahn</i>  |     |
| Promoting and Tracing High School Students’ Identity Change in an Augmented Virtual<br>Learning Environment.....               | 216 |
| <i>Hamideh Talaftian, Mamta Shah, Amanda Barany, Aroutis Foster</i>  |     |
| Collaborative Talk Across Two Pair-Programming Configurations .....  | 224 |
| <i>Zarifa Zakaria, Danielle Boulden, Jessica Vandenberg, Jennifer Tsan, Collin Lynch, Eric Wiebe, Kristy Elizabeth Boyer</i>   |     |
| CoProStory: A Tangible Programming Tool for Children’s Collaboration.....  | 232 |
| <i>Xiaozhou Deng, Danli Wang, Qiao Jin</i>   |     |
| A Spiral Model of Collaborative Knowledge Improvement to Support Collaborative .....   | 240 |
| <i>Wenli Chen, Si Zhang, Yun Wen, Chee-Kit Looi, Jennifer Yeo</i>  |     |
| Flexible CSCL Orchestration Technology: Mechanisms for Elasticity and Dynamism in<br>Pyramid Script Flows .....                | 248 |

*Kalpani Manathunga, Davinia Hernández-Leo*

|  |     |
|--|-----|
| Please Don't Shoot the Messenger! Prompts in Online Learning Groups - Influences of Nudging Messages' Sender and Publicness on Recipients' Perception and Attribution..... | 256 |
| <i>Filipa Stoyanova, Nicole Krämer</i>   |     |
| How Augmented Reality Affects Collaborative Learning of Physics: a Qualitative Analysis.....   | 264 |
| <i>Apittha Unahalekhaka, Iulian Radu, Bertrand Schneider</i>   |     |
| Developing a Model of Collaborative Learning with Minecraft for Social Studies Classrooms Using Role-play Theory and Practice.....   | 272 |
| <i>Anders Mørch, Louise Mifsud, Siv Eie</i>  |     |
| "Can't Nobody Floss Like This!": Exploring Embodied Science Learning in the Third Space.....   | 280 |
| <i>Chris Georgen</i>   |     |
| Fostering Collective Knowledge Advancement Through Idea-Friend Maps in a Large Knowledge Building Community.....   | 288 |
| <i>Xueqi Feng, Jan van Aalst, Carol Chan, Yuqin Yang</i>   |     |
| Using and Perceiving Emoji in Design Peer Feedback.....  | 296 |
| <i>Na Sun, Elise Lavoué, Chulakorn Aritajati, Aurélien Tabard, Mary Beth Rosson</i>  |     |
| On the Understanding of Students' Learning and Perceptions of Technology Integration in Low- and High-embodied Group Learning.....   | 304 |
| <i>Marianna Ioannou, Yiannis Georgiou, Andri Ioannou, Mina Johnson</i>   |     |
| Effective Regulation in Collaborative Learning: An Attempt to Determine the Fit of Regulation Challenges and Strategies.....   | 312 |
| <i>Nadine Melzner, Martin Greisel, Markus Dresel, Ingo Kollar</i>  |     |
| Unpacking Socio-Metacognitive Sense-Making Patterns to Support Collaborative Discourse.....  | 320 |
| <i>Marcela Borge, Tugce Aldemir, Yu Xia</i>  |     |
| The Binary Replicate Test: Determining the Sensitivity of CSCL Models to Coding Error.....   | 328 |
| <i>Brendan Eagan, Zachari Swiecki, Cayley Farrell, D.W. Shaffer</i>  |     |
| Developing Relationships, Changing Participation: Computational Identity.....  | 336 |
| <i>Melissa Gresalfi, Amanda Bell, Alexandria Cervantes</i>   |     |
| Family Collaboration in the Digital Age: Parent Learning Partner Roles Are Linked to Child Expertise and Parents' Work.....  | 344 |
| <i>Brigid Barron, Caitlin K. Martin, Judy Nguyen</i>   |     |
| From Material Objects to Social Objects: Researching the Material-Dialogic Spaces of Joint Attention in a School-based Makerspace.....                                     | 352 |
| <i>Kristiina Kumpulainen, Anu Kajamaa</i>  |     |
| Analyzing Students' Synergistic Learning Processes in Physics and CT by Collaborative Discourse Analysis.....  | 360 |
| <i>Caitlin Snyder, Nicole Hutchins, Gautam Biswas, Mona Emara, Shuchi Grover, Luke Conlin</i>  |     |
| Girls as Experts, Helpers, Organizers, and Leaders: Designing for Equitable Access and Participation in CSCL Environments.....   | 368 |
| <i>Kay Ramey, Reed Stevens</i>   |     |
| Reading for Breadth, Reading for Depth: Understanding the Relationship Between Reading and Complex Thinking Using Epistemic Network Analysis.....                          | 376 |
| <i>Hanall Sung, Shengyang Cao, Andrew Ruis, D.W. Shaffer</i>   |     |
| On a Making-&-Tinkering Approach to Learning Mathematics in Formal Education: Knowledge Gains, Attitudes, and 21st-Century Skills.....                                     | 384 |

*Stella Timotheou, Andri Ioannou*

|   |     |
|---|-----|
| Parent and Educator Perspectives on the Benefits of an Online Space to Promote Offline Program Collaborative Learning .....                                       | 392 |
| <i>Susan Walker, Cheuk Hei, Bosco Cheng, Alize Rattenni, Caroline Reynolds, Sam Lapham</i>  |     |
| An Alternate Statistical Lens to Look at Collaboration Data: Extreme Value Theory .....   | 400 |
| <i>Kshitij Sharma, Jennifer Olsen</i>   |     |
| Exploring Disciplinary Boundaries in Early Elementary Students' Developing Practices.....   | 408 |
| <i>Kylie Peppler, Joshua Danish, Naomi Thompson</i>   |     |
| Assessing Collaborative Problem Solving: What and How? .....  | 416 |
| <i>Haley Tsang, Seung Won Park, Laure Lu Chen, Nancy Law</i>  |     |
| Metaphorical Reasoning Together: Embodied Conceptualization in a Community of Philosophical Inquiry.....  | 424 |
| <i>Polo Claire, Sandra Lagrange-Lanaspre</i>  |     |
| Tracing Teacher Collaborative Learning and Innovation Adoption: a Case Study in an Inquiry Learning Platform.....   | 432 |
| <i>María Jesús Rodríguez-Triana, Luis P. Prieto, Tobias Ley, Ton de Jong, Denis Gillet</i>  |     |
| A Scaled Analysis of How Minecraft Gamers Leverage YouTube Comment Boxes to Participate and Collaborate .....   | 440 |
| <i>Justice Walker, Stefan Slater, Yasmin Kafai</i>  |     |
| How Teachers Implement Active Learning: Typologies of Orchestration Flow .....  | 448 |
| <i>Elizabeth Charles, Jim Slotta, Robert Cassidy, Michael Dugdale, Chao Zhang, Kevin Lenton</i>   |     |
| Cross-boundary Interaction for Sustaining Idea Development and Knowledge Building With Idea Thread Mapper.....  | 456 |
| <i>Guangji Yuan, Jianwei Zhang, Mei-Hwa Chen</i>  |     |
| Embodiment of Computational Thinking During Collaborative Robotics Activity .....   | 464 |
| <i>Theodore Kopcha, Ceren Ocak</i>  |     |
| Bugs as a Nexus for Emergent Peer Collaborations: Contextual and Classroom Supports for Solving Problems in Electronic Textiles .....                             | 472 |
| <i>Deborah Fields, Gayithri Jayathirtha, Yasmin Kafai</i>   |     |
| Supporting Collaborative Curriculum Customizations Using the Knowledge Integration Framework .....  | 480 |
| <i>Korah Wiley, Allison Bradford, Marcia Linn</i>   |     |
| The Affordable Touchy Feely Classroom: Textbooks Embedded with Manipulable Vectors and Lesson Plans Augment Imagination, Extend Teaching-Learning Practices ..... | 488 |
| <i>Durgaprasad Karnam, Harshit Agrawal, Priyanka Borar, Sanjay Chandrasekharan</i>  |     |
| Evaluating an Adaptive Equity-Oriented Pedagogy on Student Collaboration Outcomes Through Randomized Controlled Trials.....                                       | 496 |
| <i>Andrew Phuong, Judy Nguyen</i>   |     |
| The Effectiveness of Publicly vs. Privately Assigned Group Leaders Among Learners in Rural Villages in Tanzania .....   | 504 |
| <i>Judith Uchidiuno, Evelyn Yarzebinski, Emilio Vargas-Vite, Ken Koedinger, Amy Ogan</i>  |     |
| Taking the Patch Perspective: A Comparative Analysis of a Patch Based Participatory Simulation .....  | 512 |
| <i>Lauren Vogelstein, Corey Brady</i>   |     |
| Understanding the Effect of Group Variance on Learning .....  | 520 |
| <i>Sinan Yavuz, Michael Forkner, Dana Gnesdilow, Catherine Dornfeld Tissenbaum, Jee-Seon Kim, Sadhana Puntambekar</i>   |     |



|   |     |
|---|-----|
| Arguing about Synthetic Biology in 140 Characters or Less: Affordances of Microblogging for High School Students Discussions of Socioscientific Issues..... | 526 |
| <i>Mia Shaw, Justice Walker, Yasmin Kafai</i>   |     |
| Developing Productive Discourse through Collective Inquiry of Knowledge-Building Principles .....   | 534 |
| <i>Yuyao Tong and Carol Chan</i>  |     |

## Volume 2

### Short Papers

|  |     |
|--|-----|
| Co-Design for Learner Help-Giving Across Physical and Digital Contexts .....   | 545 |
| <i>Ishrat Ahmed, Victor Giroto, Areej Mawasi, Amanda Whitehurst, Ruth Wylie, Erin Walker</i>                                     |     |
| College Goes Online With Avatars: Extended Learning in a Middle School .....   | 549 |
| <i>Martine Gadille, Maria Antonietta Impedovo</i>  |     |
| Moving Between Experience, Data and Explanation: The Role of Interactive GIS Maps in Elementary Science Sensemaking .....        | 553 |
| <i>Kathryn Lanouette, Sarah Van Wart</i>   |     |
| Young Interpreters: The Situated as Supportive.....  | 557 |
| <i>Kevin Johnstun, Heather Leary, Bryant Jensen</i>  |     |
| Collaboration Script Appropriation in a Science Class .....  | 561 |
| <i>Si Zhang, Wenli Chen, Yun Wen, Chee-Kit Looi, Aileen S.W. Chai, JL Ang</i>  |     |
| Understanding the Connections of Collaborative Problem Solving Skills in a Simulation-based Task Through Network Analysis.....   | 565 |
| <i>Mengxiao Zhu, Jessica Andrews Todd</i>  |     |
| Video Annotation for Content-Focused Coaching .....  | 569 |
| <i>Julie Amador, Jeffrey Choppin, Cynthia Carson, Ryan Gillespie</i>   |     |
| Negotiating Uncertainty to Develop a Joint Deepening Focus in Knowledge Building Discourse.....                                  | 573 |
| <i>Simona Pesaresi, Jianwei Zhang</i>  |     |
| Foundations of Community in an Online, Asynchronous Professional Development Website.....  | 577 |
| <i>Genevieve Henricks, Victoria Jay, Shereen Beilstein, Michelle Perry, Meg Bates, Cheryl Moran, Joseph Cimpian</i>              |     |
| Exploring Teaching and Course Assistants' Interventions with Groups During Collaborative Problem-Solving.....                    | 581 |
| <i>Saadeddine Shehab, Emma Mercier</i>   |     |
| Students' Funds of Knowledge and Knowledge Creation During STEM Learning in a Computer-supported Makerspace .....                | 585 |
| <i>Anu Kajamaa, Kristiina Kumpulainen</i>  |     |
| Children's Interactive Strategies Around Digital Technology in Collaborative Learning Environments .....                         | 589 |
| <i>Dhvani Toprani, Mona Alqahtani, Marcela Borge</i>   |     |
| Examining the Educative Value of Person-to-person Knowledge Sharing on Social Media: The Case of YouTube as a Site of CSCL?..... | 593 |
| <i>Ilana Dubovi, Iris Tabak</i>  |     |
| Designing Epistemic Scaffolding in CSCL .....  | 597 |
| <i>Feng Lin, Sadhana Puntambekar</i>   |     |

|  |     |
|--|-----|
| Collaborative Remembering, temporal cement of Collaborative Learning: An Exploration .....   | 601 |
| <i>Lucas Bietti, Michael Baker</i>   |     |
| Social Media in the Science Classroom: Bridging Funds of Knowledge to Scientific Concepts .....  | 605 |
| <i>Kelly Mills, Elizabeth Bonsignore, Tamara Clegg, Jason Yip, June Ahn, Daniel Pauw, Caroline Pitt</i>  |     |
| An Initial Look at the Developing Culture of Online Global Meet-ups in Establishing a Collaborative,<br>STEM Media-Making Community.....                             | 608 |
| <i>Danielle Espino, Seung Lee, Brendan Eagan, Eric Hamilton</i>  |     |
| Application of the IBE-UNESCO Global Competences Framework in Assessing STEM-focused, Global<br>Collaborative Learning Within a Digital Makerspace Environment ..... | 612 |
| <i>Danielle Espino, Seung Lee, Lauren Van Tress, Eric Hamilton</i>   |     |
| SciGirls Code: Computational Participation and Computer Science with Middle School Girls .....   | 616 |
| <i>Cassie Scharber, Lana Peterson, Yu-Hui Chang, Sarah Barksdale, Ramya Sivaraj, Angelina Constantine,<br/>Jennifer Englund</i>                                      |     |
| A Systematic Review of the Quantification of Qualitative Data in Proceedings of International<br>Conferences on CSCL from 2005 to 2017.....                          | 620 |
| <i>Yu Xia, Marcela Borge</i>   |     |
| Reflective Assessment for Idea Improvement Through Collective Concept Mapping.....   | 624 |
| <i>Hyejin Park, Jianwei Zhang</i>  |     |
| Mapping Individual to Group Level Collaboration Indicators Using Speech Data .....   | 628 |
| <i>Cynthia D'Angelo, Jennifer Smith, Nonye Alozie, Andreas Tsiartas, Colleen Richey, Harry Bratt</i>   |     |
| Advancing Elementary Students' Reading Comprehension Through Knowledge Building .....  | 632 |
| <i>Pei-Yi Lin, Huang-Yao Hong, Leanne Ma</i>   |     |
| Using Digital Interrupted Case Studies for Whole Class Inquiry in Life Sciences .....  | 636 |
| <i>Leslie Schneider, Berri Jacque, Dr. Jim Slotta, Emily Kate McDonough, Jessica Henry</i>   |     |
| Longitudinal Analysis and Visualization of Participation in Online Courses Powered by Cohesion<br>Network Analysis .....   | 640 |
| <i>Maria-Dorinela Sirbu, Mihai Dascalu, Scott Crossley, Danielle McNamara, Stefan Trausan-Matu</i>   |     |
| The Impacts of Online Peer Feedback Supported by Argumentation Instruction and Argumentative<br>Scripts on Students' Learning Outcomes.....                          | 644 |
| <i>Saeed Latifi, Omid Noroozi, Javad Hatami, Harm Biemans</i>  |     |
| Embedding Microblogging Technology to Support Classroom Dialogue .....   | 648 |
| <i>Anja Amundrud, Ingvill Rasmussen</i>  |     |
| BacToMars: A Collaborative Video Game for BioDesign.....   | 652 |
| <i>Clarissa Verish, Amanda Strawhacker, Lauren Westendorf, Christina Pollalis, Amanda Sullivan, Anna<br/>Loparev, Marina Bers, Orit Shaer</i>                        |     |
| Digital Badges for Capturing, Recognizing, Endorsing, and Motivating Broad Forms of<br>Collaborative Learning .....  | 656 |
| <i>Daniel T. Hickey, Christopher Andrews, Grant Chartrand</i>  |     |
| Embodied Collaboration to Foster Instrumental Genesis in Mathematics .....   | 660 |
| <i>Anna Shvarts, Rosa Alberto, Arthur Bakker, Michiel Doorman, Paul Drijvers</i>   |     |
| Collaborative Uncertainty Management While Solving an Engineering Design Problem.....  | 664 |
| <i>Navneet Kaur, Aastha Patel, Chandan Dasgupta</i>  |     |
| Understanding Climate Change Through Collaborative Versus Individual Inquiry With Constructive or<br>Example-Based Scaffolds.....                                    | 668 |
| <i>Sarah Bichler, Sonya Richards, Lisa Hasenbein, Marcia Linn, Frank Fischer</i>   |     |

|   |     |
|---|-----|
| Emergent Roles, Collaboration, and Conceptual Outcomes for Two Eighth-Grade Groups in CSCL Science Classes.....                           | 672 |
| <i>Catherine Dornfeld Tissenbaum, Dana Gnesdilow, Nicole Martin, Sadhana Puntambekar</i>  |     |
| Examining How Scientific Modeling Emerges Through Collective Embodied Play.....   | 676 |
| <i>Xintian Tu, Joshua Danish, Chris Georgen, Megan Humburg, Bria Davis, Noel Enyedy</i>   |     |
| See the Collaboration Through the Code: Using Data Mining and CORDTRA Graphs to Analyze Blocks-Based Programming.....                     | 680 |
| <i>Mike Tissenbaum, Vishesh Kumar</i>   |     |
| Towards Automatic and Pervasive Physiological Sensing of Collaborative Learning.....  | 684 |
| <i>Kshitij Sharma, Ilias Pappas, Sofia Papavlasopoulou, Michail Giannakos</i>   |     |
| Scaffolding Inclusivity Through Making: A Preliminary Analysis of Diverse Learners' Meaning Making Through Designing Complex Systems..... | 688 |
| <i>Gabriela Richard, Sagun Giri, Robert Ashley, Nakisha Whittington</i>   |     |
| Exploring Group' Design Thinking Patterns in a Principe-based Knowledge-Building Environment.....   | 692 |
| <i>Pei-Yi Lin, Huang-Yao Hong</i>   |     |
| Perspective Taking in Participatory Simulation-based Collaborative Learning.....  | 696 |
| <i>Fengfeng Ke, Xin Yuan, Mariya Pachman, Zhaihuan Dai, Raymond Naglieri, Xinhao Xu</i>   |     |
| Using A Virtual Design Studio to Support Collaborative Studio Instruction.....  | 700 |
| <i>Jessica Briskin, Susan Land</i>  |     |
| Maggie Mars: Theatrical Modeling and the Phenomenological Understanding of Solar Systems.....   | 704 |
| <i>Jackson Reimers, Corey Brady</i>   |     |
| Impact of Choice on Students' Use of an Experimentation Model for Investigating Ideas About Thermodynamics.....                           | 708 |
| <i>Jennifer King Chen, Marcia Linn</i>  |     |
| What Can Be Learned About Computer-Supported Collaborative Learning From a Bibliographic Coupling Analysis?.....                          | 712 |
| <i>Heisawn Jeong, Sujong Seo, Jiyeon Jeong, Cindy Hmelo-Silver, Sebastian Grauwin</i>   |     |

## Symposia

|   |     |
|---|-----|
| Group Formation in the Digital Age: Relevant Characteristics, Their Diagnosis, and Combination for Productive Collaboration.....  | 719 |
| <i>Dimitra Tsovaltzi, Armin Weinberger, Lara Schmitt, Henrik Bellhäuser, Adrienne Müller, Johannes Konert, René Röpke, hannie gijlers, Elise Eshuis, Tessa Eysink, Melanie Erkens, Sven Manske, H. Ulrich Hoppe, Daniel Bodemer, Sreecharan Sankaranarayanan, Cameron Dashti, Christopher Bogart, Xu Wang, Majd Sakr, Michael Hilton, Carolyn Rosé</i>                                  |     |
| Ghost in the Machine: A Symposium on Collaboration Between Human and Computerized Agents in Educational Contexts.....   | 727 |
| <i>Matthias Stadler, Frank Fischer, Art Graesser, Pantelis M. Papadopoulos, Anika Radkowsch, Ralf Schmidmaier, Martin Fischer, Haiying Li, Jiangang Hao, Samuel Greiff</i>  |     |
| A Wide Lens on Learning in a Networked Society: What Can We Learn by Synthesizing Multiple Research Perspectives?.....  | 735 |
| <i>Yael Kali, Iris Tabak, Dani Ben-Zvi, Christopher Hoadley, Hava Ben-Horin, Dror Angel, Keren Aridor, Osnat Atias, Ayelet Baram-Tsabari, Maya Benichou, Oren Golan, Yaela N Golumbic, Yotam Hod, Carmel Kent, Adi Kidron, Hana Manor, Nakhi Mishol-Shauli, Carmit Pion, Sheizaf Rafaeli, Amit Rechavi, Malka Shacham, Amit Schejter, Noam Tirosh, Patrice L. Weiss, Oren Zuckerman</i> |     |
| Computerized Text Analysis: Assessment and Research Potentials for Promoting Learning.....  | 743 |
| <i>Hee-Sun Lee, Danielle McNamara, Zoë Buck Bracey, Ou Lydia Liu, Libby Gerard, Bruce Sherin, Chris Wilson, Amy Pallant, Marcia Linn, Kevin C. Haudek, Jonathan Osborne</i>   |     |

|   |     |
|---|-----|
| Technology-Mediated Teacher-Researcher Collaborations: Professional Learning Through Co-Design.....   | 751 |
| <i>Susan Goldman, Andrea Gomoll, Cindy Hmelo-Silver, Allison Hall, Monlin Ko, Angela Fortune, Eleni Kyza, Andria Agesilaou, Kimberley Gomez, Louis Gomez, Emily Pressman, Katherine Rodela, Iris Tabak</i>  |     |
| Tools and Methods for ‘4E Analysis’: New Lenses for Analyzing Interaction in CSCL.....  | 759 |
| <i>Rolf Steier, Ben Shapiro, Dimitra Christidou, Palmyre Pierroux, Jacob Davidsen, Rogers Hall</i>  |     |
| The Roles of Knowledge in Knowledge Creation .....  | 767 |
| <i>Carl Bereiter, Carol Chan, Huang-Yao Hong, Jiyeon Lee, Ahmad Khanlari, Pei-Yi Lin, Ching Sing Chai, Chin-Chung Tsai, Marlene Scardamalia, Seng Chee Tan, Yuyao Tong, Jan van Aalst, Jianwei Zhang, Yibing Zhang</i>  |     |
| Theorizing and Measuring Collective Productive Disciplinary Engagement.....   | 775 |
| <i>Britte Cheng, Melissa Gresalfi, Amanda Bell, Corey Brady, Lauren Vogelstein, Crina Damsa, Tuire Palonen, Toni Rogat, Anne Traynor, Temitope Adeoye, Andrea Gomoll, Cindy Hmelo-Silver, Dan Hickey, Christopher Andrews, Grant Chartrand, Rebecca Itow, Patrik Lundh, Sten Ludvigsen</i>  |     |
| Understanding CSCL Through the Lens of Research Syntheses.....  | 783 |
| <i>Freydis Vogel, Heisawn Jeong, Susan Yoon, Stian Håklev, Leonore Valentine Guillain, Nour Ghalia Abassi, Sally Wai Yan Wan, Sancia Wai-San Wan, Anika Radkowsch, Frank Fischer, Cindy Hmelo-Silver</i>  |     |
| Designing for Productive Problem Posing in Informal STEM Spaces .....   | 791 |
| <i>Katherine Chapman, Lara Jasien, Paul Reimer, Lauren Vogelstein</i>   |     |
| Personal Embodiment, Social Enactment: Collaborative Learning With Body Technology .....  | 799 |
| <i>Deborah Fields, Cynthia Carter Ching, Victor Lee, Breanne Litts, Chase Mortensen, Joshua Danish, Megan Humburg, Bria Davis, Xintian Tu, Debora Lui, Mia Shaw, Gayithri Jayathirtha, Yasmin Kafai, Lindsay Lindberg, Robb Lindgren, James Planey, Jason Morphew, Katie Headrick Taylor Taylor, Adam Bell, Erin Riesland, Maria Hays, Deborah Silvis, Ilana Dubovi, Noel Enyedy, Rogers Hall</i> |     |
| <b>Posters</b>  |     |
| Effects of a Digital Guided Peer Feedback System on Student Learning and Satisfaction .....   | 809 |
| <i>Omid Noroozi, Arash Bayat, Javad Hatami</i>  |     |
| StandUp: Engaging Professionals to Coach Design Projects .....  | 811 |
| <i>Daniel Rees Lewis, Matthew Easterday, Elizabeth Gerber</i>   |     |
| Scaling Dialogic Peer Feedback via Learning Analytics and Scripts .....   | 813 |
| <i>Erkan Er, Yannis Dimitriadis, Dragan Gasevic</i>   |     |
| A Case Study on the Development of Pre-Service Teachers’ Design Thinking.....   | 815 |
| <i>Chih Hui Seet, Huang-Yao Hong, Ching Sing Chai</i>   |     |
| Rural Students' Cultural Assets During Science Argumentation.....   | 817 |
| <i>Jingjing Sun, Brian W. Miller, Michael Finlay</i>  |     |
| Assessing Iterative Planning for Real-world Design Teams .....  | 819 |
| <i>Daniel Rees Lewis, Elizabeth Gerber, Matthew Easterday</i>   |     |
| Perceptions of Online Professional Development: Do Newer and Experienced Teachers Differ?.....  | 821 |
| <i>Denise Bressler, Susan Yoon, Katherine Miller, Joeeun Shim, Daniel Wendel, Ilana Schoenfeld, Emma Anderson, David Reider</i>   |     |
| Leveraging the Power of Collaborative Reflection to Promote Learning Through Clinical Practice.....   | 823 |
| <i>Rupal Jain, Meredith Moore</i>   |     |
| Development of Girls’ Interests and Identities in Computer Sciences Within an CSCL Environment.....   | 825 |
| <i>Yu-Hui Chang, Sarah Barksdale, Lana Peterson, Cassie Scharber</i>  |     |
| Using Participatory Design to Facilitate In-service Teacher Learning of Computational Thinking.....   | 827 |
| <i>Shamya Karumbaiah, Sugat Dabholkar, Joeeun Shim, Susan Yoon, Betty Chandy, Andy Ye</i>   |     |

|   |     |
|---|-----|
| A Review of the Evolving Definition of Orchestration: Implications for Research and Design .....  | 829 |
| <i>LuEttaMae Lawrence, Emma Mercier</i>   |     |
| Conceptualizing and Analyzing the Instructor-Student Collaboration .....  | 831 |
| <i>Fan Ouyang</i>   |     |
| Conflicts and Collaboration: A Study of Upper Elementary Students Solving Computer Science Problems ....  | 833 |
| <i>Jennifer Tsan, Jessica Vandenberg, Xiaoting Fu, Jamieka Wilkinson, Danielle Boulden, Kristy Elizabeth Boyer, Collin Lynch, Eric Wiebe</i>        |     |
| Social Media and Personal Histories: Practices, Identities, and Algorithms in the Age of Digital Nostalgia ....                                     | 835 |
| <i>Sarah Burriss</i>  |     |
| Initial Knowledge and the Intensity of Online Discussion .....  | 837 |
| <i>Miiikka Turkkila, Henri Lommi, Ismo Koponen</i>  |     |
| An Innovative Social-Cognitive Engagement Network Representation.....   | 839 |
| <i>Fan Ouyang, Yu-Hui Chang</i>   |     |
| Reinforce Context Awareness in Augmented Reality-based Learning Design .....  | 841 |
| <i>Yun Wen, Chee-Kit Looi</i>   |     |
| Citizen Science in Schools: Supporting Implementation of Innovative Learning Environments Using Design-Centric Research-Practice Partnerships ..... | 843 |
| <i>Maya Benichou, Osnat Atias, Ornit Sagy, Yael Kali, Ayelet Baram-Tsabari</i>  |     |
| Designing for ESM-Mediated Collaborative Science Learning .....   | 845 |
| <i>Sugat Dabholkar, Uri Wilensky</i>  |     |
| Contribution to the Integration of MOOC in a Hybrid-Learning Project in the Moroccan University .....   | 847 |
| <i>Bouchaib Riyami</i>  |     |
| From Face-to-Face to Online: Considerations for a Social Capital Design to Scale Science Teacher Access to High-Quality PD .....                    | 849 |
| <i>Susan Yoon, Kate Miller, Joeeun Shim, Daniel Wendel, Ilana Schoenfeld, Emma Anderson, David Reider</i>   |     |
| Epistemic Frames of Idea Evaluation in Collaboration .....  | 851 |
| <i>Shono Ikeda, Jun Oshima, Ritsuko Oshima</i>  |     |
| Collaborative Knowledge Construction Mediated by Technology.....  | 853 |
| <i>Afaf Baabdullah</i>  |     |
| Online Mob Programming: Bridging the 21st Century Workplace and the Classroom .....   | 855 |
| <i>Sreecharan Sankaranarayanan, Xu Wang, Cameron Dashti, Haokang An, Clarence Ngoh, Michael Hilton, Majd Sakr, Carolyn Rosé</i>                     |     |
| Moving Beyond the “Façade of Participation”: Using Choice-based Design to Enhance Online Discussions .....  | 857 |
| <i>Valerie Barbaro</i>  |     |
| Making the Design of CSCL Analytics Interfaces a Co-design Process: the Case of Multimodal Teamwork in Healthcare.....                              | 859 |
| <i>Roberto Martinez-Maldonado, Vanessa Echeverria, Doug Elliott, Carmen Axisa, Tamara Power, Simon Buckingham Shum</i>                              |     |
| Examining the Role of Emotion Awareness and Sharing Emotions during Collaborative Learning .....  | 861 |
| <i>Maedeh Kazemitabar, Elise Lavoué, Susanne Lajoie, Rubiella Carrillo, Tenzin Doleck</i>   |     |
| Computer-based Problem Solving to Prepare for Adaptive Consolidation .....  | 863 |
| <i>Antje Boomgaarden, Katharina Loibl, Timo Leuders</i>   |     |
| The Relationship Between Young Students’ Attitudes Toward Collaboration and Team Satisfaction in a STEAM-based Program .....                        | 865 |



*Tuba Ketenci, Betul Erkan*

|  |     |
|--|-----|
| Theorizing and Operationalizing Social Engagement as a Precursor to Productive Disciplinary Engagement in Collaborative Groups.....        | 867 |
| <i>Toni Kempler Rogat, Britte Cheng, Andrea Gomoll, Temitope Adeoye, Anne Traynor, Cindy Hmelo-Silver, Patrik Lundh</i>                    |     |
| Embedding Computational Thinking in the Elementary Classroom: An Extended Collaborative Teacher Learning Experience.....                   | 869 |
| <i>Diane Jass Ketelhut, Emily Hestness, Kelly Mills</i>  |     |
| Visualizing Representations of Interaction States during CSCL .....  | 871 |
| <i>Saadeddine Shehab, Emma Mercier</i>   |     |
| Exploration of Scaffolding in Teachers' Dialogue Analysis .....  | 873 |
| <i>Moegi Saito, Shinya Ikubo, Hajime Shirouzu</i>  |     |
| How Embodied Interactions Manifest Themselves During Collaborative Learning in Classroom Settings .....                                    | 875 |
| <i>Zuzanna Lechelt, Yvonne Rogers, Nicolai Marquardt</i>   |     |
| Enhancing Free-text Interactions in a Communication Skills Learning Environment .....  | 877 |
| <i>Raja Lala, Marcell van Geest, Stefan Ruseti, Johan Jeurig, Mihai Dascalu, Jordy van Dortmund, Gabriel Gutu-Robu, Michiel Hulsbergen</i> |     |
| Collaboration Within Mathland: What Do We Become Together.....   | 879 |
| <i>Isaac Nichols-Paez, Corey Brady</i>   |     |
| Using Epistemic Network Analysis to Explore Ways of Contributing to Knowledge Building Discourse.....                                      | 881 |
| <i>Leanne Ma, Yoshiaki Matsuzawa, Marlene Scardamalia</i>  |     |
| Identifying Learning Leaders in Collaborative Learning .....   | 883 |
| <i>Yinying Wang, Min Kyu Kim</i>   |     |
| Supporting Meaningful Revision of Scientific Ideas in an Online Genetics Unit .....  | 885 |
| <i>Emily Harrison, Libby Gerard, Marcia Linn</i>   |     |
| Use of Spatial Sensemaking Practices in Spatial Learning.....  | 887 |
| <i>Abha Vaishampayan, Julia Plummer, Patricia Udomprasert, Susan Sunbury</i>   |     |
| Teacher Perceptions on Collaborative Online Professional Development for In-Service Teachers on a MOOC Platform .....                      | 889 |
| <i>Katherine Miller, Susan Yoon, Jooeun Shim, Daniel Wendel, Ilana Schoenfeld, Emma Anderson, David Reider</i>                             |     |
| Shall We Learn Together in Loud Spaces? Towards Understanding the Effects of Sound in Collaborative Learning .....                         | 891 |
| <i>Milica Vujovic, Davinia Hernández-Leo</i>   |     |
| Assessing Collaborative Problem Solving in the Context of a Game-based Learning Environment .....  | 893 |
| <i>Asmalina Saleh, Yuxin Chen, Abeera Rehmat, Karyn Housh, Cindy Hmelo-Silver, Krista Glazewski, James Lester</i>                          |     |
| Structure for Agency: Possibilities and Challenges for Adaptive Collaborative Learning Support in Educational Equity Projects.....         | 895 |
| <i>Suraj Utamchandani, Cindy Hmelo-Silver</i>  |     |
| Conceptualizations of Learning in ijCSCL.....  | 897 |
| <i>Jonan Donaldson</i>   |     |
| The Role of Asynchronous Digital Feedback in Youth Maker Projects.....   | 899 |
| <i>Erica Halverson, Amy Mueller, Zhaohui Dai</i>   |     |

|  |     |
|--|-----|
| Using Collaborative Agent-based Modeling to Explore Complex Phenomena with Elementary Preservice Science Teachers .....  | 901 |
| <i>Tony Petrosino, Maximilan Sherard, Corey Brady</i>  |     |
| Robots to Help Us Feel Safe: A Problem-Based CSCL Experience .....   | 903 |
| <i>Andrea Gomoll</i>   |     |
| An Exploration of Female Engagement and Collaboration in the Bricks and Bits Maker Project .....   | 905 |
| <i>Ezequiel Aleman, Pryce Davis</i>  |     |
| The Development and Application of a Social Reading Platform and the Double-level Scaffolding .....  | 907 |
| <i>Xiangdong Chen, Chen Guo</i>  |     |
| Imagination in Adolescents' Collaborative Multimodal Science Fictions.....   | 909 |
| <i>Maria Kolovou, Ji Shen, Blaine Smith</i>  |     |
| Telepresence Robots as Embodied Agents in the Classroom .....  | 911 |
| <i>Penny Thompson, Yam Chaivisit</i>   |     |
| Child-Material Computing: Material Collaboration in Fiber Crafts .....   | 913 |
| <i>Anna Keune, Kylie Peppler</i>   |     |
| Dialogic Intervisualizing: Rethinking Text-Discourse-Learning Relations in Multimodal Problem-based Learning.....  | 915 |
| <i>Susan Bridges, Lap Ki Chan, Judith Green, Asmalina Saleh, Cindy Hmelo-Silver</i>  |     |
| Exploring the Potential of IVR Technology to Promote Collaborative Learning in Science Experiences.....  | 917 |
| <i>Omar Ceja-Salgado, Sara Price</i>   |     |
| Using Computer-Supported Collaborative Learning (CSCL) for Global Curriculum Inquiry: A Case Study in Hong Kong .....  | 921 |
| <i>Sally Wai Yan Wan</i>   |     |
| A Qualitative Analysis of Joint Visual Attention and Collaboration with High- and Low-Achieving Groups in Computer-Mediated Learning.....  | 923 |
| <i>Tonya Bryant, Iulian Radu, Bertrand Schneider</i>   |     |
| An Interactive Visualization to Enhance Social Learning Practices in MOOC Platforms .....  | 925 |
| <i>Philip Tubman, Phil Benachour, Murat Oztok</i>  |     |
| Revision Analysis of Students' Position-Time Graphs.....   | 927 |
| <i>Emily Harrison, Elizabeth McBride, Marcia Linn</i>  |     |
| Research as Learning From Youth: Leveraging Collaborative Digital Tools to Position Youth as Experts on Themselves .....   | 929 |
| <i>Cynthia Graville, Joseph Polman, Taylor Morgan, Claire Englander, Jordan Fair, Kurt Lott, Tessa McGartland, BriYana Merrill, Kennedy Morganfield, Darby Moore, Annie O'Brien, Adam Rush, Patrick Shanahan, Erik Swenson, Ben Sylar, Michael Teasedale, Erikah White</i> |     |
| Mezclado: History of Our Neighborhoods in Augmented Reality .....  | 931 |
| <i>Kit Martin</i>  |     |
| Studying Computational Thinking Practices Through Collaborative Design Activities with Scratch .....   | 933 |
| <i>Joey Huang, Kylie Peppler</i>   |     |
| Computer-Supported Collaborative Learning and Learning Sciences Research in EducMap.....   | 935 |
| <i>Heisawn Jeong, Joomi Kim, Kristine Lund, Sebastian Grauwin</i>  |     |
| Using a Resource Activation Lens to Understand Classroom Enactments of Computationally-Based Science Curricula.....  | 937 |
| <i>Joeun Shim, Susan Yoon, Noora Noushad</i>   |     |
| Assessing Collaborative Online Inquiry and Social Deliberation in Digital Environments .....   | 939 |

*Jesse R. Sparks, Julie Coiro, Jill Castek, Carita Kiili, Beth Holland, Changhee Lee*

|  |     |
|--|-----|
| Learning with Multiple Representations and Student Engagement in Secondary Education: A Preliminary Review of Literature .....                                   | 941 |
| <i>Engida Gebre, Addisu Bailie</i>   |     |
| An Exploratory Study of Automated Clustering of Themes to Identify Conceptual Threads in Knowledge Building Discourse .....                                      | 943 |
| <i>Gaoxia Zhu, Leanne Ma, Andrew Toulis, Monica Resendes</i>   |     |
| Exploring Students' Self-assessment on Collaborative Process, Calibration, and Metacognition in an Online Discussion.....  | 945 |
| <i>Yu Xia, Hyeyeon Lee, Marcela Borge</i>  |     |
| Intelligent Cognitive Assistants to Support Orchestration in CSCL.....   | 947 |
| <i>Haesol Bae, Krista Glazewski, Cindy Hmelo-Silver, James Lester, Bradford W. Mott, Jonathan Rowe</i>   |     |
| Mediating Collaboration in History with Network Analysis .....   | 949 |
| <i>Haesol Bae, Kalani Craig, Joshua Danish, Cindy Hmelo-Silver, Suraj Uttamchandani, Maksymilian Szostalo</i>  |     |
| Impacts to Students of Infused Computational-Thinking, Problem-Based-Learning Science Curriculum: Engagement, Opportunity to Learn, Interest, and Identity ..... | 951 |
| <i>Amy Semerjian, Collette Roberto</i>   |     |
| When Words Are Not Enough: What Student Gestures and Embodied Responses Tell Us About Understanding Science Through Dance .....                                  | 953 |
| <i>Lindsay Lindberg, Danielle Keifert, Noel Enyedy, Joshua Danish</i>  |     |
| Using Sentence Embeddings to Automatically Extract Cohesion and Alignment Metrics in Problem-Solving Tasks.....  | 955 |
| <i>Alejandro Andrade, Chris Georgen, Michael Stucker</i>   |     |
| Gaming the Schoolyard: Promoting High School Students' Collaborative Learning Through Geolocate Mobile Game Design.....  | 957 |
| <i>Judy Perry</i>  |     |
| Designing Learning Analytics for Teacher Learning: An Analytics-Supported Teacher Professional Development (ASTPD) Approach.....                                 | 959 |
| <i>Gaowei Chen, Kennedy Chan, Carol Chan, Jinjian Yu, Hu Liru, Jiajun Wu, Lauren Resnick</i>   |     |
| Balancing the Scales: Implications of Model Size for Mathematical Engagement .....   | 961 |
| <i>Kylie Peppler, R. Mishael Sedas, Anna Keune, Suraj Uttamchandani</i>  |     |
| Knowledge Building, Robotics, and Math Education.....  | 963 |
| <i>Ahmad Khanlari, Marlene Scardamalia</i>   |     |
| Microblogging for Joint Construction of Meaning in the Classroom .....   | 965 |
| <i>Jan Arild Dolonen, Ingvill Rasmussen, Sten Ludvigsen</i>  |     |
| Tools to Facilitate Teacher and Student Collaboration in Assessment.....   | 967 |
| <i>Kate Thompson, Harry Kanasa, Susan Chapman</i>  |     |
| Negotiation of Epistemic Territories and Collaborative Learning in Workplace Interactions: The Case of Requests for Assistance .....                             | 969 |
| <i>Vasiliki Markaki-Lothe, Laurent Filliettaz</i>  |     |
| Understanding Teachers' Collaboration for Designing Technology-Enhanced Learning.....  | 971 |
| <i>Chunli Wang, Xiaoqing Gu</i>  |     |

## **Special Sessions**

|   |     |
|---|-----|
| Eight Provocations for the CSCL Field: Current Significant Discussions and the Future of CSCL ..... | 975 |
|---|-----|

*Alyssa Wise, Baruch Schwarz, Sten Ludvigsen, Nikol Rummel, Yotam Hod, Ornit Sagy, Yael Kali, Marcela Borge, Emma Mercier, Pierre Tchounikine*

|   |     |
|---|-----|
| IAALDE Best Paper Session and Panel Discussion .....  | 977 |
| <i>Rémi Venant, Kshitij Sharma, Pierre Dillenbourg, Phillipe Vidal, Julien Broisin, Jennifer Olsen, Vincent Aleven, Nikol Rummel, Alyssa Wise, Carolyn Penstein Rosé, Gaëlle Molinari, Danielle McNamara, Wenli Chen, Stephanie Teasley</i> |     |
| What Can 4E Cognition Tell Us About the Successes and Failures of CSCL Implementation .....   | 978 |
| <i>Claire Polo, Andri Ioannou, Didier Bottineau, Jean-Michel Roy, Magali Ollagnier Beldame</i>  |     |

## **Interactive Events**

|  |      |
|--|------|
| RoomCast: Distributing Digital Resources in the Classroom of Things .....  | 981  |
| <i>Tom Moher, Alessandro Gnoli</i>   |      |
| Computational Action in App Inventor: Developing Theoretical and Technological Frameworks for Collaboration and Empowerment .....  | 985  |
| <i>Mike Tissenbaum, Josh Sheldon</i>   |      |
| Toccatà: A Multi-Device System for Activity Scripting and Classroom Orchestration .....  | 989  |
| <i>Valentin Lachand, Aurélien Tabard, Christine Michel</i>   |      |
| Bridging Concepts as Intermediary Knowledge in Design: Productive Dialogues and the Talkwall Microblogging Tool.....               | 993  |
| <i>Ole Smørdal, Ingvill Rasmussen</i>  |      |
| ElectroVR: An Electrostatic Playground for Collaborative, Simulation-Based Exploratory Learning in Immersive Virtual Reality ..... | 997  |
| <i>Scott W. Greenwald, Wiley Corning, Gavin McDowell, Pattie Maes, John Belcher</i>  |      |
| The CUBE: A Tangible for Embodied Learning, Balanced Engagement, and Classroom Orchestration .....                                 | 1001 |
| <i>Pantelis M. Papadopoulos</i>  |      |
| Synergy: An Online Platform for Dialogic Peer Feedback at Scale .....  | 1005 |
| <i>Erkan Er, Yannis Dimitriadis, Dragan Gasevic</i>  |      |
| Lines We Trace: Comparing Data Displays to Support Youth Sailing .....   | 1009 |
| <i>Katherine Chapman, Ben Rydal Shapiro</i>  |      |
| FROG, A Tool to Author and Run Orchestration Graphs: Affordances and Tensions.....   | 1013 |
| <i>Stian Håklef, Louis Faucon, Jennifer Olsen, Pierre Dillenbourg</i>  |      |
| Mission HydroSci: Meeting Learning Standards Through Gameplay.....   | 1017 |
| <i>James Laffey, Joseph Griffin, Justin Sigoloff, Troy Sadler, Sean Goggins, Andrew Womack, Eric Wulff, Sean Lander</i>            |      |
| Project Bloks: Embodied and Collaborative Learning With Tangible Interfaces for Young Children .....                               | 1021 |
| <i>Veronica Lin, Paulo Blikstein</i>   |      |
| The Power of Network Analysis Tool for Collaborative Learning.....   | 1025 |
| <i>Haesol Bae, Kalani Craig, Joshua Danish, Cindy E. Hmelo-Silver, Suraj Uttamchandani, Maksymilian Szostalo, Ann McCranie</i>     |      |
| Supporting Collaborative Problem Solving in a Game-Based Learning Environment.....   | 1029 |
| <i>Asmalina Saleh, Yuxin Chen, Cindy Hmelo-Silver, Krista Glazewski, Bradford Mott, Robert Taylor, Jonathan Rowe, James Lester</i> |      |

## **Workshops**

|  |      |
|--|------|
| EPCAL: Studying Synchronous Online Collaboration at Scale .....                            | 1035 |
| <i>Jiangang Hao, Diego Zapata-Rivera, Jessica Andrews-Todd, Mengxiao Zhu, Piotr Mitros</i> |      |

|  |      |
|--|------|
| Experiencing and Analysing Embodied Design Thinking in CSCL.....   | 1036 |
| <i>Penny Wheeler, Stefan Trausan-Matu, Amanda Barany, Jonan Phillip Donaldson</i>  |      |
| Theories and Methods for Researching Interdisciplinary Learning.....   | 1037 |
| <i>Lina Markauskaite, Hanni Muukkonen, Crina Damsa, Peter Reimann, David Williamson Shaffer, Kate Thompson, Yael Kali, Adi Kidron, Monika Nerland</i>                                    |      |
| Designing Embedded Phenomena .....   | 1038 |
| <i>Tom Moher, James Slotta, Brenda Lopez Silva, Michelle Lui, Beth Sanzenbacher</i>  |      |
| Posthumanist Perspectives on Learning .....  | 1039 |
| <i>Kylie Peppler, Anna Keune, Karen Wohlwend, Jennifer Rowsell, Robert Goldstone</i>   |      |
| Theorizing and Analyzing Productive Disciplinary Engagement as a Collaborative Phenomenon .....  | 1040 |
| <i>Toni Kempler Rogat, Britte Haugan Cheng</i>   |      |
| Making the Learning Sciences Count: Impacting Association for Computing Machinery Communities in Human-Computer Interaction.....   | 1041 |
| <i>Jason C. Yip, June Ahn, Elizabeth Bonsignore, Tamara Clegg, Elizabeth DiSalvo</i>   |      |
| Creating, Refining, and Validating Automated Discourse Codes: An Introduction to nCoder and Rho.....   | 1042 |
| <i>Amanda Siebert-Evenstone, Brendan Eagan, Zachari Swiecki, Seung Bok Lee, Eric Hamilton</i>  |      |
| Creating Common Ground for Teaching Designs in Learning Sciences Programs: Building on CSCL Research and Practice.....   | 1043 |
| <i>Freydis Vogel, Marcela Borge, Crina Damsa, Pryce Davis, Julia Eberle, Frank Fischer, Daniel Sommerhoff</i>  |      |
| International Workshop on Collaboration Analytics: Making Learning Visible in Collaborative Settings.....  | 1044 |
| <i>Roberto Martinez Maldonado, Marcelo Worsley, Bertrand Schneider, Ahmed Kharrufa</i>   |      |
| Understanding New Assessment and Environment for Knowledge Building: Triangulating Features of Discourse Platform, Multimodal Learning Analytics and Text-based Learning Analytics ..... | 1045 |
| <i>Chew Lee Teo, Elizabeth Koh, Marlene Scardamalia, Jianwei Zhang, Mei-Hwa Chen, Carol Chan, Sambit Praharaj, Esther Tan, Aloysius Ong</i>  |      |
| Publishing in the Learning Sciences: A Journal Writers' Workshop .....   | 1047 |
| <i>Mitchell J. Nathan, Carol Chan, Sten Runar Ludvigsen, Joshua Radinsky, Jeremy Roschelle, Susan Yoon, Jan van Aalst</i>  |      |
| Mid-Career Workshop: Empowering Women in the Learning Sciences.....  | 1048 |
| <i>Stephanie D. Teasley, Carol Chan, Susan R. Goldman, Kimberley Gomez, Janet L. Kolodner, Marcia C. Linn, Nichole Pinkard, Nikol Rummel</i>   |      |
| <b>Early Career Workshop</b>   |      |
| CSCL 2017 Early Career Workshop .....  | 1051 |
| <i>Julia Eberle, Nancy Law, Victor R. Lee</i>  |      |
| An Embodied Conjecture Approach Towards Designing Games Using Problem-based Learning.....  | 1052 |
| <i>Asmalina Saleh</i>  |      |
| Supporting Youth in Exploring and Expressing Disciplinary Identities With Learning Technologies .....  | 1054 |
| <i>Shiyan Jiang</i>  |      |
| Cultivating an Orientation to Care in Computing Education.....   | 1056 |
| <i>Ben Rydal Shapiro</i>   |      |
| Talking to Learn Across Digital and Face-to-face Settings in a High School English Classroom.....  | 1058 |
| <i>Jennifer Higgs</i>  |      |
| Improving Feedback for Learning Real-world Design.....   | 1060 |

*Daniel Rees Lewis*

|   |      |
|---|------|
| Supporting Student Understanding of Engineering: Using Hands-on Projects and Interactive Computer Models .....          | 1062 |
| <i>Elizabeth McBride</i>  |      |
| Devising an Interactive Social Learning Analytics Tool to Foster Online Collaborative Learning .....                    | 1064 |
| <i>Fan Ouyang</i>   |      |
| Combining Machine Learning and Learning Analytics to Provide Personalized, Adaptive Scaffolding .....                   | 1066 |
| <i>Irene-Angelica Chounta</i>   |      |
| Towards the Development of an Evidence-based Framework for Immersion in Digitally Enhanced Learning Environments .....  | 1068 |
| <i>Yiannis Georgiou</i>   |      |
| Conceptualizing and Scaling Dialogic Peer Feedback .....  | 1070 |
| <i>Erkan Er</i>   |      |
| Learner Modeling of Cognitive and Metacognitive Processes for Complex Collaborative Learning Tasks .....                | 1072 |
| <i>Shitanshu Mishra</i>   |      |
| Towards Personally-Relevant Learning: Bridging In-School and Out-of-School Learning Through Wearable Technologies ..... | 1074 |
| <i>Sharon Lynn Chu</i>  |      |
| Flexible Scripting and Orchestration for Synchronous Learning.....  | 1076 |
| <i>Stian Håklev</i>   |      |

## **Doctoral Consortium**

|   |      |
|---|------|
| The CSCL 2017 Doctoral Consortium Workshop .....  | 1079 |
| <i>Yotam Hod, Katerine Bielaczyc, Heisawn Jeong, Joshua Danish, Mimi Recker</i>   |      |
| Learning in the City: Joint Pursuits in Relational Pivot Points.....  | 1082 |
| <i>Adam Bell</i>  |      |
| Teacher Identity and the Co-Design of CSCL That Creates Change.....   | 1084 |
| <i>Andrea Gomoll</i>  |      |
| Modeling and Modes: Broadening Participation in Science Practices for Emerging Bilingual Students.....  | 1086 |
| <i>Ashlyn Pierson</i>   |      |
| Sustaining Knowledge Building Across Communities With Boundary Objects.....   | 1088 |
| <i>Guangji Yuan, Jianwei Zhang</i>  |      |
| Emergent Leadership in Student-led Collaborative Activity in a School-based Makerspace.....   | 1090 |
| <i>Jasmiina Korhonen</i>  |      |
| Temporal Patterns and Visualizations of Peer Talk: Toward Understanding the Process and Performance of Dialogic Collaborative Problem-Solving ..... | 1092 |
| <i>Liru Hu, Gaowei Chen, Carol K.K. Chan</i>  |      |
| Interplay of Different Group Awareness Information for Improving Collaborative Learning in Social Media .....                                       | 1094 |
| <i>Lisa Ollesch, Daniel Bodemer</i>   |      |
| Weaving Together: Exploring How Pluralistic Mathematical Practices Emerge Through Weaving .....   | 1096 |
| <i>Naomi Thompson</i>   |      |
| Implicit Guidance, Explicit Guidance, or Both? Examining the Differential Effects of Collaboration  |      |

|  |      |
|--|------|
| Prompts and Group Awareness Tools on Students' Regulation of Online Collaboration .....                                    | 1098 |
| <i>Sebastian Strauß</i>  |      |
| Towards Equitable Learning Futures: Sociopolitical Discourses, Practices, and Joint Work in an<br>LGBTQ+ Youth Group ..... | 1100 |
| <i>Suraj Uttamchandani</i>   |      |
| Modeling the Contributions of Individuals to Collaborative Problem Solving Using<br>Epistemic Network Analysis .....       | 1102 |
| <i>Zachari Swiecki, David Williamson Shaffer</i>   |      |

## **Indexes**

|                     |       |
|---------------------|-------|
| Author Index.....   | A1-A7 |
| Keyword Index ..... | K1-K3 |

# **Keynote Presentations**





# Integrating Theorizing on Embodied, Enactive, Extended, and Embedded Cognition to Augment CSCL Research

Stephen Fiore, University of Central Florida, [sfiore@ist.ucf.edu](mailto:sfiore@ist.ucf.edu)

**Abstract:** The metaphor of the blind men and the elephant is often used to describe how one's narrow perspective on an issue leads to misperceptions about some complex issue. Interdisciplinary research is no stranger to this parable; so much so that it has become clichéd. With the advent of the embodied, enactive, extended, and embedded cognition perspectives, we arrive at a place where we must avoid misconceptions about learning and cognition and discuss how to address these developments in the context of CSCL. Said most succinctly, through the embodiment thesis, one considers cognition as inextricably linked to the sensory-motor system with which one has been endowed. Via the enactive thesis, one considers cognition as tightly coupled between interactor and the environment. With the extended thesis, one considers cognition as existing as much in the world as it is in the head. In the embedded thesis, one considers cognition as best understood when situated within a particular context. On the one hand, like the blind men and the elephant, we could embrace these fractionalizations and pursue independent lines of inquiry to understand how technology can support distinct forms of learning and cognition. On the other hand, we could pursue a more holistic approach that integrates these varied theoretical perspectives to consider how technologies can augment the study of learning and cognition. In this talk I discuss how new tools and technologies allow us to embrace the latter approach via a multi-method, multi-conceptual approach with the group as the focal point. This is made possible through advances in interdisciplinary research allowing us to instrument and/or observe the world of interaction in ways never before possible. Importantly, though, we are observing interaction not just within, but also across, multiple levels. I discuss how this provides the opportunity to integrate levels of interaction for studying learning and cognition and provide representative research issues for CSCL to explore.

*Dr. Stephen M. Fiore* is Director of the Cognitive Sciences Laboratory, and Professor with the University of Central Florida's Cognitive Sciences Program in the Department of Philosophy and Institute for Simulation & Training. He is Past-President of the Interdisciplinary Network for Group Research and a founding committee member for the annual Science of Team Science Conference. He maintains a multidisciplinary research interest that incorporates aspects of the cognitive, social, organizational, and computational sciences in the investigation of learning and performance in individuals and teams. His primary area of research is the interdisciplinary study of complex collaborative cognition and the understanding of how humans interact socially and with technology.

# Negotiating Knowledge, Expertise, Connoisseurship and Taste in Social Interaction

Lorenza Mondada, University of Basel, [lorenza.mondada@unibas.ch](mailto:lorenza.mondada@unibas.ch)

**Abstract:** How knowledge is manifested in social interaction is a key issue for understanding and documenting how knowledge is possibly negotiated, recognized, and learned in actual settings and activities. Recent work in interactional studies has debated how knowledge in interaction can be not only claimed but also displayed by self, how it can be attributed to self by others, how it can be intersubjectively recognized and negotiated (Heritage, 2012, Stivers, Mondada, Steensig, 2011). This might be observed in ordinary life, as well as in institutional contexts: in the latter case, institutional asymmetries might enhance epistemic asymmetries – for instance between lay persons and experts – but also be the arena for epistemic re-negotiations and competitions – for instance between amateurs, connoisseurs and professionals. These recalibrations of knowledge claims and attributions are central in settings that are not devoted to formal learning but that are relevant for socialization into knowledge, culture, and expertise. Empirically the talk focuses on an exemplary setting of ordinary life in which these issues are observable and documentable in actual video recordings of situated practices. Encounters in specialized gourmet shops, in particular in cheese shops, constitute an arena in which participants engage not only in buying, but in displaying their identity as gourmet connoisseurs; an arena in which clients go not just for shopping but for learning how to taste and also how to speak about sophisticated food items. Knowledge, expertise and taste concern in this case both propositional knowledge (information about the products) and praxeological and embodied knowledge (know how to taste and to assess the products). The paper discusses different types of knowledge, verbal and embodied, and the way they are embedded in specific activities and sequential contexts. It offers methodological hints about their analysis, providing for both situated and systematic patterns, showing how they are expressed and enacted in situ, in negotiations that crucially involve talk and bodies, constructing social and epistemic identities.

*Professor Lorenza Mondada* holds an appointed Chair of General and French Linguistics at the University of Basel. She is a leading worldwide expert in the field of linguistics, who has published widely on the interrelationship of language, embodiment, space and mobility. In 2001 she was awarded the National Latsis Prize – a prize awarded annually in Switzerland to the top researcher under 40. Her empirical studies of multimodal interaction in a wide range of activities – from rallying to cheese tasting – have been instrumental in viewing human language as a dynamic resource. Since January 2015 she is a Finnish Distinguished Professor at the University of Helsinki, where she develops a project entitled “Revisiting Language through Multimodality in Interaction”.

# Designing Opie Robots as Learning Companions: Insights, Interactions and Interdependencies

Janet Wiles, University of Queensland, [j.wiles@uq.edu.au](mailto:j.wiles@uq.edu.au)

**Abstract:** This talk will present a series of case studies in technology design, based on the child-friendly robots in the ongoing Opal project and analysis tools for interaction dynamics. Opie robots are primarily designed for physically-embodied social presence, enabling children and adults to touch, hold, or hug the robot, and to use its solid frame as a physical support. The key design issues start with safety of the users, which affects movement and speed; and safety of the robot from rough play by children. These capabilities enable different kinds of studies to those of commonly available commercial robots which are too fragile for rough use, or in danger of falling over, with potential for damage to themselves and a danger to children. The second design consideration for Opie robots is the speed of interaction, and how that can impact on human social engagement, which occurs in timescales of hundreds of milliseconds. Case studies will include Opie robots as story tellers in public spaces, such as science museums and technology showcases; Indigenous language robots as language assistants in classrooms and language centres; Lingodroids that evolve their own languages; and chatbots used for surveys and conversation. Insights from the development of multi-lingual robots include the critical role of embedding robots in communities and the extended nature of the robots' influence within and beyond a classroom. The Opie real-world case studies enable us to reflect on fundamental questions about design decisions that affect when a robot is considered to be a social being; whether a robot could understand the grounded meanings of the words it uses; and practical questions about what is needed for a robot to be an effective learning companion in individual and group settings. The talk will conclude with an overview of two computational tools for analysing human-human and human-robot interactions, including conceptual recurrence analysis of turn-taking in conversations using Discursis and the timing of interactions using Calpy's pausecode.

*Dr. Janet Wiles* is Professor at the School of Information Technology and Electrical Engineering (ITEE) in the Faculty of Engineering, Architecture and Information Technology, with the University of Queensland. She is research leader of the Co-Innovation group in UQ's ITEE School, and UQ node leader for the ARC Centre of Excellence for the Dynamics of Language. Her current research projects focus on Human-robot interactions, Indigenous language technologies, bio-inspired computation, visualisation and artificial intelligence. Previous projects have been in complex systems modeling in biology and neuroscience, human memory, language and cognition.



# **Full Papers**



# Analysis of Touchscreen Interactive Gestures During Embodied Cognition in Collaborative Tabletop Science Learning Experiences

Nikita Soni, Alice Darrow, Annie Luc, Schuyler Gleaves, Carrie Schuman, Hannah Neff, Peter Chang, Brittani Kirkland, Amanda Morales, Kathryn A. Stofer, Lisa Anthony  
nsoni2@ufl.edu, aldarrow@ufl.edu, annieluc@ufl.edu, schuylergleaves@ufl.edu,  
carrie.schuman@ufl.edu, hannahneff@ufl.edu, pchang27@ufl.edu, bgkirkland@ufl.edu,  
mondily@ufl.edu, stofer@ufl.edu, lanthony@cise.ufl.edu  
University of Florida

Jeremy Alexandre (9<sup>th</sup> author)  
jeremy.alexandr@gmail.com  
City University of New York—Brooklyn College

**Abstract:** Previous work has used embodied cognition as a theoretical framework to inform the design of large touchscreen interfaces for learning. We seek to understand how specific gestural interactions may be tied to particular instances of learning supported by embodiment. To help us investigate this question, we built a tabletop prototype that facilitates collaborative science learning from data visualizations and used this prototype as a testbed in a laboratory study with 11 family groups. We present an analysis of the types of gestural interactions that accompanied embodied cognition (as revealed by users' language) while learners interacted with our prototype. Our preliminary findings indicate a positive role of *cooperative (multi-user)* gestures in supporting scientific discussion and collaborative meaning-making during embodied cognition. Our next steps are to continue our analysis to identify additional touchscreen interaction design guidelines for learning technologies, so that designers can capitalize on the affordances of embodied cognition in these contexts.

## Introduction

Large touchscreen interfaces like multi-touch tabletops have become increasingly widespread, particularly in informal learning environments such as science museums (Geller, 2006). For example, science educators and researchers are using large touchscreen interfaces to support learning about complex global phenomena such as Earth's ocean system (Cheek, 2010). Yet, many visitors have trouble understanding the information presented on these interfaces, let alone interacting with the interfaces to develop a deeper understanding of the information (Cheek, 2010). Previous interaction design research has used *embodied cognition* as an underlying theoretical framework to inform the design of large touchscreen interfaces for learning (Lin et al., 2016; Piper et al., 2012). Embodied cognition theory posits that cognition is not solely based in the mind, but also in the body, i.e., some of our cognitive processes occur through "perceptually-guided motions" (Wilson & Golonka, 2013). Learning around a shared interactive display like a touchscreen tabletop with direct-touch gestures draws on the affordances of embodiment. These interactive gestures play an important role in how learners explain abstract ideas and promote scientific discussion (Piper et al., 2012). Although embodiment theories consider sensorimotor activities like direct-touch gestures to be integral to learning, the community has noted that there is still no "conceptually coherent and empirically validated design framework" to inform the design of embodied learning experiences (Abrahamson et al., 2018, p.1243). Prior research that has explored the design of touchscreen interfaces for learning from the viewpoint of embodied cognition has assumed that embodied cognition is driving learning during any motion or interaction, without considering how *specific* interactions may influence particular instances of learning. If we could identify the types of touchscreen interactions being used when people are engaged in embodied cognition during a learning episode, we could design learning experiences that explicitly support and encourage touchscreen interactions directly linked to learning.

In this paper, we analyze touchscreen gestural interactions that accompany linguistic cues of embodied cognition during collaborative learning (Kirschner et al., 2018), in the context of science learning about data visualizations of Earth's global ocean system. We built a tabletop application prototype to support collaborative learning and used this prototype as a testbed in a lab study with 11 family groups. Throughout this paper, we use the term "learning" to reflect *meaning-making*, that is, the process of integrating new knowledge and coordinating it with existing beliefs and knowledge (Vygotsky, 1978), rather than its traditional use to reflect "knowledge acquisition." Based on previous research that has shown conceptual metaphors are a type of embodiment language (Lakoff & Johnson, 2003), we relied on identifying the three main conceptual metaphors



(*metonymy*, *orientation*, and *ontology*) in groups' utterances as cues to the occurrence of embodied cognition. We then analyzed participants' gestures that co-occurred with these utterances and linked these gestures to how people were making meaning at that moment. Specifically, we investigated the following research questions: (RQ1) When groups' utterances reveal that embodied cognition is occurring, what gestures are they making, during learning with an interactive tabletop? (RQ2) How can we support and encourage these types of gestural interactions in the design of computer-supported collaborative learning experiences on an interactive tabletop, to afford embodied cognition more directly? Our preliminary findings indicate a positive role of *cooperative (multi-user)* touchscreen gestures in supporting collaborative learning on tabletops during embodied cognition. This paper outlines two initial themes and touchscreen interaction design guidelines for affording embodied cognition, to inform the design of future tabletop experiences for science learning.

## Embodied cognition in designing tabletop applications for learning

Prior work investigating collaborative learning around multi-touch tabletops has already used embodied cognition as a theoretical framework to design these learning experiences (Lin et al., 2016; Piper et al., 2012). Embodied cognition provides an ideal framework for understanding interactions with multi-touch technology because it considers direct hands-on interactions with the digital content to be integral to cognition (Kirsh, 2013). Lin et al. (2016) used embodied cognition as their theoretical basis to design a tabletop application to promote collaborative thinking among high school students. Their findings show that externalizing learners' thinking by adding scaffolding such as sticky notes made it easier for groups to share and reflect on their thoughts collaboratively. They note that being able to enlarge or zoom sticky notes facilitated joint reading and group thinking, implying that touchscreen gestures and collaborative thinking are intertwined. Piper and Hollan (2009) compared affordances of tabletops and pen-paper material for supporting scientific discussions with undergraduate students and noted that tabletops better supported understanding of abstract science concepts, such as how a neuron fires. Learners used bimanual and collaborative whole-handed gestures over various parts of the axon during scientific discourse. On the other side, prior work has also explored how to design direct-touch interaction generally to support science learning activities around tabletops (Horn et al., 2009; Shaer et al., 2011), although not explicitly from the standpoint of embodied cognition. Horn et al. (2009) developed an information visualization tool (*Involv*) for exploring the Encyclopedia of Life to help groups of adults effectively interact with large datasets on the tabletop. Shaer et al. (2011) explored tabletop interactions for classroom science learning and found that multi-touch tabletops encourage reflection and foster collaboration. However, these studies focus more broadly on the role of interaction in supporting science learning, rather than looking at *specific* interactions that augment scientific discussion and collaborative learning. Our work builds upon both of these lines of prior work. In addition to using embodied cognition as our conceptual framework, we analyze the gestures that are co-occurring with instances of embodied cognition (as signaled by the users' language) to understand the role of gestures in facilitating the collaborative science learning process.

## Language as demonstration of embodied cognition

The theory of embodied cognition posits that learning occurs not only within the mind, but also in the body, through the learner's movements and interactions with the environment (Wilson & Golonka, 2013). Lakoff and Johnson (2003) discussed language as one signal of embodiment, specifically the use of "conceptual metaphors" for space and place, outlining three types (Table 1). *Oriental metaphors* use spatial words for ideas that are not inherently spatial. *Ontological metaphors* allow an abstract idea, such as an experience, to represent a concrete substance like an object. A specific example is *personification*: giving objects human-like qualities so we can relate experiences through human characteristics. *Metonymy* is when one idea stands in for another similar idea, as when someone speaks about a location they are not physically in as "here". Wilson and Golonka (2013) have critiqued some embodied cognition research, however, suggesting that linguistic expression of metaphors found in much embodied cognition research (including this paper) is at best a precursor to true embodied cognition, as language still occurs in the mind. The tasks investigated so far that Wilson and Golonka

Table 1: Examples of the three types of conceptual metaphors from Lakoff and Johnson (2003).

| Conceptual Metaphors | Examples  |
|----------------------|---|
| <b>Oriental</b>      | When we use "up" to describe feeling happy, e.g., "I'm feeling <i>up</i> ." (p.16)  |
| <b>Ontology</b>      | When a physical object is described as "talking" or "giving" as if it were a person, e.g., "His religion <i>tells</i> him that he cannot drink fine French wines." (p.28)   |
| <b>Metonymy</b>      | When we use the name of one location to describe our experience associated with the events at that physical location, such as using the term "Grand Central Station" to mean "a crowded place," e.g., "It's been <i>Grand Central Station</i> here all day." (p.32) |

suggest represent “true” embodied cognition tasks, such as catching a fly baseball, are considerably less “representation hungry” than our more abstract task of learning from data visualizations. Our task deals with thinking about data that is not concretely present in the learner’s environment (Kiverstein & Rietveld, 2018). For examine, one part of our task is understanding what the colors in the data visualizations depict, another part is determining the pattern(s) of the colors, and a third part is ascribing meaning to the patterns. Decomposing our meaning-making task into elements that engage embodied cognition as Wilson and Golonka do is a question outside the scope of this paper. The specific conceptual metaphors learners use are likely to be influenced by the specific science domain; however, our focus is not to categorize the types of conceptual metaphors used by learners in this domain. Instead, we focus on linking the touchscreen gestures during embodied cognition to the learning that is occurring. These insights will inform the design of future touch-interactive learning experiences to more successfully capitalize on the affordances of embodied cognition in data visualization tasks.

## Our prototype application for learning about Earth’s ocean system

To study our research questions, our team, which consisted of both human-computer interaction experts and learning scientists, designed a touch-interactive tabletop prototype for science data visualizations to engage family groups in collaborative meaning-making. The application ran on a Samsung SUR40 tabletop computer. The resolution was 1920 x 1080 (55 DPI), and the display size was 40 inches, measured diagonally. We created the interface with the Open Exhibits Software Development Kit (SDK) (1).

To aid groups in the collaborative meaning-making process, we provided scaffolding (Stofer, 2016), e.g., cognitive affordances that can help non-expert users understand the ocean temperature data visualizations better, such as audience-appropriate color schemes and geographic labels. We used two map views with color

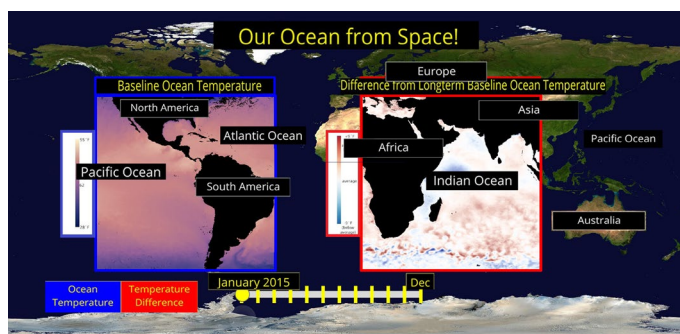


Figure 1. Prototype design with the red/blue maskviewers on a base Earth map, along with a time slider (yellow bar) to change months. Text labels have been enlarged for legibility.

scaffolding designed by NASA (2). The visualizations we used are similar to the NASA Earth Observations visualizations, which have been deployed in nearly 100 museums worldwide (3). Our prototype consisted of two “maskviewer” interface elements, each containing a set of ocean data visualizations from the year 2015 on top of a base Earth map with land maps and blue oceans (4) (Figure 1). The first visualization used a pink-to-purple color scale that showed baseline ocean temperatures, and the second visualization used a red-to-blue color scale to represent the extremes of the temperature difference (e.g., anomaly) from baseline (Stofer, 2016). The maskviewers implemented “layers” of information which could be zoomed, rotated, or dragged to allow users to flexibly control them and facilitate shared communication. To help participants recognize the difference between the two maskviewers, we attached a temperature legend and title directly to the maskviewers. Each visualization contained six continent hotspots that, when tapped, would pop-up an information box with content about El Niño, an ocean phenomenon affected by temperature, which could then be zoomed, rotated, or dragged in relation to that continent. We chose the content for these information boxes from international weather and climate organizations (5) to supplement the content shown on the data visualizations. Finally, to enable participants to explore how baseline temperatures or temperature anomalies changed over a year, our prototype contained a “time slider” (yellow bar in Figure 1). When the user performed a hold (long tap) anywhere on the map, the time slider would pop up, and the user could slide their finger horizontally along it to change the months and observe the continuous temperature changes. The time slider had tick marks for each month, labels at either end of the slider, and the current month displayed above the user’s finger. We used this prototype as a testbed to identify interactive touchscreen gestures that co-occurred with instances of embodied cognition while groups engaged in a scientific discussion about ocean temperatures.

## Participants and study design

A total of 30 participants in 11 family groups participated in our study (16 female). Each group consisted of at least one child (ages 8 to 13, M: 10.07 yrs, SD: 1.49 yrs, 15 children total) and one parent or guardian (max group size: four). We recruited participants via an email sent to a faculty list and flyers distributed at a local science museum. Our protocol was approved by our Institutional Review Board. After obtaining informed consent and assent, we instructed groups in how to use the think-aloud process (Greenberg et al., 2011) to

expose what they were thinking while completing the learning activity. We also did a practice think-aloud with them as a group to solve a two-column addition problem. During the study, we asked each group to perform these four tasks in order, while thinking aloud: (1) Explore this interactive visualization as you would if you saw it in a science museum. We are interested in seeing what people do. Tell us what you find out about the ocean. (2) Open the information box for South America and tell us how the information there compares to the ocean data displayed in general. (3) Find the Gulf of Mexico and tell us how ocean temperatures there change month to month. (4) Find the Eastern Pacific Ocean basin and tell us how ocean temperatures in the basin compare to long-term baseline ocean temperatures for that basin.

The order of tasks increased in difficulty, with the aim that, as participants explored the prototype, they would be better able to answer the more involved questions. After the tasks, participants completed a demographics questionnaire. Most of our participants (over 85%) were frequent science museum visitors (e.g., three to twelve times per year). Each family group received a \$30 grocery store gift card for participating.

## Data collection and analysis

During the study, we video recorded participants using cameras placed at a side angle and directly overhead. Session lengths (excluding break time between tasks) ranged from 14 minutes (min) 30 seconds (sec) to 27 min 45 sec (average duration: 21 min 43 sec, just under 6 min per task). We transcribed the groups' utterances from the videos. One group spoke some Spanish during their interaction, and we asked a colleague to assist with translation. To understand how specific interactions may be tied to particular instances of learning supported by embodiment, we first identified the set of learning "episodes" in which groups' utterances signaled embodied cognition (e.g., from the presence of conceptual metaphors). As mentioned, the three main types of conceptual metaphors are *metonymy*, *orientation*, and *ontology* (Lakoff & Johnson, 2003). Three researchers reviewed the transcripts to identify utterances containing these conceptual metaphors. We then analyzed the participants' touchscreen gestures (e.g., tap, pinch to zoom, etc.) that co-occurred with these utterances during each learning episode, to help us understand the role of these gestures in supporting collaborative making-meaning, as facilitated by embodied cognition. Our team conducted a thematic analysis on these examples, in which we discussed themes that emerged in terms of which types of gestures most often accompanied the conceptual metaphors and seemed to be essential in affording embodied cognition during collaborative meaning-making.

## Findings

We present two over-arching themes that illustrate the types of gestures groups used to facilitate collaborative meaning-making when engaged in embodied cognition in our study.

### Gestures for orienting the group

We observed that participants' interactions with the interface elements guided their thinking process and helped them orient themselves and the group to the science content displayed on the prototype. We saw that the physical affordances offered by the maskviewer element, such as being able to resize and move it around the tabletop, allowed participants to focus the group's attention on how temperatures change at various geographic locations. Segment 1 (Figure 2) is an example of this theme. We identified it as an instance of embodied cognition based on lines 4, 5, and 6. These utterances illustrate the use of *metonymic* conceptual metaphor, in which P2 and P3 [Group597] are speaking as if placing themselves physically into the world represented by the prototype: they are talking about the temperature change as if it was actually occurring, saying that "*It's really cold!*" and using words "*...out here in like the middle...*" even though P2 is not in the middle of the ocean himself. This segment shows how participants are dragging the maskviewer element to incrementally build their understanding of the temperature visualization by promoting discussion about temperature variations at specific geographic locations. In their gestures, P2 is dragging the baseline ocean temperature maskviewer element over the Indian Ocean, as he says, "*So check that out, that's very cold right there.*" P2 further drags the maskviewer near the middle of the Pacific Ocean to compare the temperature patterns for the Indian and Pacific Oceans, as he says, "*So out here in like the middle of the Pacific Ocean, it looks like it's upwards of 88 degrees.*"

We also observed that resizing the maskviewer helped focus the group's attention and facilitate collaborative meaning-making. Segment 2 (Figure 2) is an example of this theme. We identified it as an instance of embodied cognition based on line 2, which exemplifies a common type of ontological metaphor called personification: P2 [Group912] treats the prototype as a "storyteller" by saying that the prototype is "telling" them something. In their gestures, we see the group resizing the maskviewer to focus the group's attention and facilitate meaning-making. Initially, P2 [Group 912] directs P1 to re-size the maskviewer in order to focus on a specific geographic location, as he says: "*Let me see Florida ... Ya, shrink it and put it like right over the north east ... So, it tells what different temperatures are here or ...*". At this moment, P1 is using the maskviewer

element as a tool to help direct the group’s attention onto Florida and analyze how the temperature changes near it: “*Pretty warm.*” The above examples show dragging and zooming gestures co-occurring with the utterances signifying embodied cognition. Thus, we can infer that participants were using the maskviewer element as an embodied lens for comparing temperature patterns through physical interactions with the prototype. Without the maskviewer, the data visualization would encompass the entire interface, making it more difficult to direct the group’s attention and focus on subcomponents of the dataset before the group is ready to think at a higher level. Dragging and zooming the maskviewer and attending to what temperature patterns they reveal for different geographical locations seemed to push participants to focus on a subset of the data at a time as their conceptual understanding is constructed piece by piece. Therefore, future touchscreen interfaces for learning should support interactions that bring key aspects of the science content gradually into focus in this way.

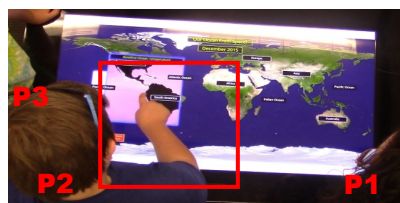
### Cooperative gestures for facilitating collaborative meaning-making

Another theme frequent in our analysis was participants using cooperative (multi-user) gestures to facilitate collaborative meaning-making. Morris et al. (2006) defined *cooperative gestures* as interactions in which the interface interprets the simultaneous gestures of more than one user as contributing to a single command. We observed that the interaction constraints offered by our time slider, in that it required simultaneous gestures (either using two hands or by two users at the same time), encouraged participants to actively participate in collaborative meaning-making through cooperative gestures. Segment 3 (Figure 3) is an example of this theme. We identified it as an instance of embodied cognition based on lines 5 and 6, which illustrates the use of *metonymic* conceptual metaphor: saying “...go to...” and “...go ahead...” to stand in for manipulating time within the visualization, as one cannot physically move to the time under discussion. This segment presents an example of how interface elements that allow users to interact cooperatively encourage group members to actively participate and contribute to the group’s shared knowledge. In this example, initially, we see that P3 is passively watching P1 and P2 [Group247] interact with the prototype and discuss their observations, instead of actively participating and contributing to the group’s understanding. We see later on (starting line 6) that since P1 is utilizing both of his hands (right hand for holding the slider and left hand to point towards the temperature pattern on the map), he looks at P3 to ask for help with changing months using the time slider by saying: “*Go to like, go to like May or June. That’s April. Look how much warmer it’s getting. That’s June. Look how super warm it is.*” Though P3 was just helping P1 change months, directly manipulating the prototype content helps P3 to focus on the temperature trends P1 and P2 were discussing. While P3 changes the month on the slider, he

#### Segment 1

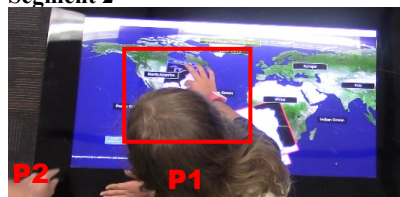


1. P1: Yeah, yeah let’s do the Indian Ocean (*dragging the maskviewer to the Indian Ocean*).
2. P2: So, check that out, that’s very cold right there\* (*dragging the maskviewer towards Europe and the top of Africa*)
3. P1: Wooooow!



4. P2: Up in here it’s pretty chilly too\*<sup>†</sup> (*dragging the maskviewer towards the Pacific Ocean*).
5. P3: It’s really cold<sup>†</sup>
6. P2: Find a really hot spot. So out here in like the middle of the Pacific Ocean, it looks like it’s upwards of 88 degrees<sup>†</sup> [Group 597]

#### Segment 2



1. P1: Let me see Florida (*dragging maskview over Florida, then enlarging maskview*)  
... (other talk)
2. P2: Ya, shrink it (P1 shrinks maskviewer) and put it like right over the north east\* (P1 drags the maskviewer over Florida). Here you go. So, it tells what different temperatures are here<sup>‡</sup> or ... (P1 zooms in map over Florida)
3. P1: Pretty warm [Group 912]

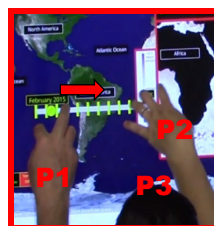
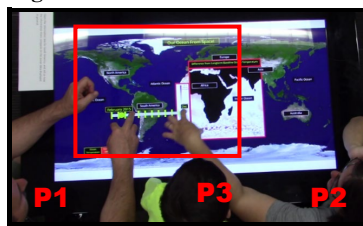
Figure 2. Example of a family group using the maskviewer to compare temperature patterns for geographic locations. The zoomed in view of the image on the right shows details of participants’ interaction with the prototype. Participants’ utterances signifying conceptual metaphors are noted: *metonymic*<sup>†</sup>, *personification*<sup>‡</sup>. Utterances marked with an asterisk (\*) co-occurred with the gestures shown in the images.



starts contributing to the group’s understanding by making inferences about how temperature patterns change at different geographical locations, as he says: “*But that’s like in the middle. Here, let me show you.*” Gestures can serve as a mechanism for cognitive offloading; that is, by taking up some of the cognitive efforts of attention and focus, they allow a learner to focus cognitive resources on other aspects of a task such as drawing inferences (Goldin-Meadow & Beilock, 2010). Furthermore, when working collaboratively on a complex task, reducing individual cognitive load and encouraging group members to exchange knowledge helps in collaborative knowledge-building (Kirschner et al., 2018). In their gestures, the participants work together in order to operate the time slider. P2 holds his hand on the screen in order for the time slider to pop up, while P3 drags the slider to different months. Thus, instead of having one participant focus their cognitive resources on operating the time slider, the group collaboration reduces the amount of cognitive effort required by each participant. Then, the participants can use more of their cognitive resources to engage in science discussion, seen in the utterance by P3, “*It’s [temperature] near av- [average] that’s where it was near average.*” The above example illustrates that using cooperative gestures facilitates collaborative meaning-making by encouraging group members to engage in a scientific discussion together in a more hands-on and active manner.

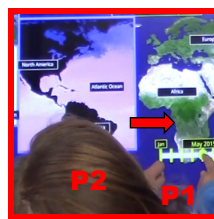
We also observed that interacting collaboratively allows group members to build upon each other’s understanding of the temperature patterns. Segment 4 (Figure 3) is an example of this theme. We identified it as an instance of embodied cognition based on lines 2, 4, 5, 6, and 8, which illustrate the use of *metonymic* conceptual metaphor: participants talk as if they are actually experiencing the time and temperature change portrayed within the prototype, by saying, “*it’s almost white now*”. In this example, we see that P1 and P2 [Group765] are collaboratively interacting with the time slider. As P2 changes months, she says, “*It’s definitely warm. June, warmer.*” P1 further builds upon the group’s understanding of temperature by saying, “*Definitely warm, it’s getting closer to 95 [degrees].*” The group continues to change months on the time slider and watch the temperature change within the visualization. Finally, P1 says, “*it’s almost white now.*” Segment 3 demonstrates that participants are using cooperative gestures to discover the mapping between different scientific variables of the visualizations (i.e., what temperature the colors depict and how temperature changes with time). Both of these examples show cooperative gestures co-occurring with utterances signifying embodied

### Segment 3



1. P2: No, that’s a different. So wait no no no. So, oh yeah temperature wouldn’t, this isn’t precipitation.
2. P1: Well yeah, but I mean they are definitely related\*. The issue that I see- (*long tap gesture to activate time slider, then dragging time slider with middle finger*)
3. P2: So warm and humid? So, if we move this over here (*dragging the maskviewer*).
4. P1: Yeah.
5. P2: And go to, it was May, June, and July right? † (*deactivates time slider by stopping long tap gesture*) That it said El Nino was? ‡
6. P1: Was it? Okay. So, if we go till, yeah go ahead\* (*long tap gesture to activate time slider*) and January the temperature isn’t that much of a difference.† Go to like, go to like May or June (*P3 dragging time slider*). That’s April. Look how much warmer it’s getting. That’s June. Look how super warm it is. †
7. P2: Yeah but you would think if it’s warm up there (*pointing gesture*) then that would cause rain over here because of warm humid air. †
8. P1: Depends on the wind pattern.
9. P3: But that’s like in the middle (*pointing gesture*). Here, let me show you. (*deactivates time slider by stopping long tap gesture*) (other talk)
10. P1: July through October
11. P3: It’s near av- that’s where it was near average. [Group 247]

### Segment 4



1. P2: Alright, what happens next? April.
2. P1: It’s getting warmer\* † (*dragging the time slider*) ...
3. P2: It’s definitely warm. June, warmer...
4. P1: Definitely warm, it’s getting closer to ninety-five. †
5. P2: Yeah that’s pretty warm. †
6. P1: Oh! Dang, that’s getting warmer. †
7. P2: July’s pretty hot.
8. P1: It’s almost white now † [Group765]

Figure 3. Examples of family groups using cooperative gestures to construct collective working memory. The zoomed in view of the image on the right shows details of participants’ interaction with the prototype. Participants’ utterances signifying conceptual metaphors are noted: *metonymic*†, *personification*‡. Utterances marked with an asterisk (\*) co-occurred with the gestures shown in the images.

cognition. Thus, we can infer that participants are using these cooperative gestures as a mechanism for cognitive offloading, and to focus their cognitive resources on meaning-making and drawing inferences related to how both time and space affect ocean temperatures. Without the cooperative gesture required by the time slider, participants would not need to coordinate to change the time period shown, making it more difficult to keep the group's understanding aligned. Using cooperative gestures seemed to push participants to create a collective working memory and encouraged group members to exchange knowledge. Therefore, future interactive learning experiences should encourage interactions that afford group members to cooperatively manipulate the interface elements like the time slider to facilitate embodied-cognition-supported learning.

## Discussion, implications, and conclusion

Prior studies offer evidence in favor of using touchscreen interaction to encourage embodied-cognition-supported learning activities around tabletops, e.g., (Lin et al., 2016). However, these studies did not analyze what specific touchscreen gestures people make when engaged in embodied cognition, and whether these gestures are tied to particular instances of learning. We seek to understand how to support and encourage the types of gestures that accompany embodied cognition as revealed in learners' language. Based on Lakoff & Johnson (2003), in our study, we used linguistic cues from groups' utterances, in the form of *conceptual metaphors*, as signals that embodied cognition was occurring. We saw instances of embodied cognition reflected in groups' language in the use of both *metonymy* and *ontology* (i.e., *personification*): by speaking as if placing themselves physically into the world represented by the prototype and by speaking as if viewing the prototype as a "storyteller". During these types of utterances, groups made a variety of touchscreen gestures, including attention-focusing gestures and cooperative gestures. Groups used the maskviewer elements as an *embodied lens* to focus the group's attention on local areas of interest and guide their thinking process, by moving and re-sizing the maskviewers to reveal the underlying temperature patterns across different geographical locations. These interactions with the maskviewer pushed participants to focus on subsets of the dataset at a time as their conceptual understanding is constructed piece by piece. In the future, designers of such learning interfaces could consider ways to incorporate similar embodied lenses (not limited to maskviewers specifically) or even multiple embodied lenses within an application that bring key aspects of the science content into focus. These lenses could be manipulated by users to explore and compare multiple regions of multidimensional data visualizations through operations such as dragging and zooming (instead of just exposing the entire dataset at once) to facilitate scientific discussion. Also, cooperative touchscreen gestures add value to applications by increasing participation and enhancing the social aspects of an interactive experience (Morris et al., 2006). Our findings indicate a positive role of such cooperative gestures in supporting scientific discussion and collaborative meaning-making, from the viewpoint of embodied cognition. We saw that using cooperative gestures supports groups in creating a collective working memory by directing the group's attention towards the same focus point and encouraging group members to contribute to the meaning-making process. Thus, we believe our findings add to the field's understanding of the role of cooperative gestures in terms of showing both how they are involved in embodied cognition and how they support collaborative meaning-making. Future tabletop learning environments could encourage similar interactions that afford group members to cooperatively manipulate the interface elements to foster group learning, as in the case of multiple lenses above.

The instances of embodied cognition we identified in participants' language suggest that their interactions with the prototype and its data were immersive and sensory-based, as participants envisioned themselves as if they were physically "in" our prototype. This finding is consistent with theories of embodiment suggesting that learners' thoughts and understandings are shaped by their prior and ongoing physical interactions with the environment (Wilson & Golonka, 2013). The language metaphors and gestures we identified support future work into examining how embodied cognition underlies the interpretation of scientific content and processes in touchscreen-based learning environments. We disagree with Goldinger et al.'s (2016) critiques of embodiment utility as a theory. Those authors do not discuss learning, and we believe the value of being able to design affordances to support learning through embodiment will be strong. Our study points toward a need for a deeper investigation of designing with embodied cognition in mind, especially in the context of science content and practices. Identifying how embodied cognition actually manifests in both language and gestures in these types of experiences will allow us to design touchscreen interactions more targeted towards fostering embodied-cognition-supported learning. In our future work, we will continue to analyze our dataset for further themes, by considering other theories related to the role of both direct touch and in-air hand gestures in facilitating collaborative meaning-making (Goodwin, 2007). While our study presented a new view of designing tabletop interfaces by identifying the nature of gestures that occur during embodied cognition, future work should evaluate the effect of encouraging these gestures to determine how well they support learners'

underlying cognitive processes in a real learning environment and expand this inquiry into other learning domains beyond data visualization and Earth's ocean temperature phenomena.

## Endnotes

- (1) <http://openexhibits.org/downloads/sdk>
- (2) <https://neo.sci.gsfc.nasa.gov/>
- (3) [https://sos.noaa.gov/What\\_is\\_SOS/sites.php](https://sos.noaa.gov/What_is_SOS/sites.php)
- (4) <https://svs.gsfc.nasa.gov/2915>
- (5) <http://www.bom.gov.au/>, <http://www.grida.no/resources/7045>

## References

- Abrahamson, D., Andrade, A., Bakker, A., Nathan, M. J., Walkington, C., Lindgren, R., ... Lindwall, O. (2018). Moving Forward: In Search of Synergy Across Diverse Views on the Role of Physical Movement in Design for STEM Education. In *International Society of the Learning Sciences* (pp. 1243–1250).
- Cheek, K. A. (2010). Commentary: A Summary and Analysis of Twenty-Seven Years of Geoscience Conceptions Research. *Journal of Geoscience Education*, 58(3), 122–134.
- Geller, T. (2006). Interactive Tabletop Exhibits in Museums & Galleries. *IEEE COMPUT GRAPH*, 26(5), 6–11.
- Goldin-Meadow, S., & Beilock, S. L. (2010). Action's Influence on Thought: The Case of Gesture. *PERSPECT PSYCHOL SCI*, 5(6), 664–674.
- Goldinger, S. D., Papesh, M. H., Barnhart, A. S., Hansen, W. A., & Hout, M. C. (2016). The Poverty of Embodied Cognition. *Psychonomic Bulletin and Review*, 23(4), 959–978.
- Goodwin, C. (2007). Environmentally Coupled Gestures. In *Gesture and the Dynamic Dimensions of Language* (pp. 195–212). John Benjamins Publishing.
- Greenberg, S., Carpendale, S., Marquardt, N., & Buxton, B. (2011). *Sketching User Experiences: The Workbook*. Elsevier.
- Horn, M., Tobiasz, M., & Shen, C. (2009). Visualizing Biodiversity with Voronoi Treemaps. In *Proceedings of the International Symposium on Voronoi Diagrams* (pp. 265–270).
- Kirschner, P. A., Sweller, J., Kirschner, F., & Zambrano R., J. (2018). From Cognitive Load Theory to Collaborative Cognitive Load Theory. *INT J COMP-SUPP COLL*, 13(2), 213–233.
- Kirsh, D. (2013). Embodied Cognition and the Magical Future of Interaction Design. *ACM Transactions on Computer-Human Interaction (TOCHI'13)*, 20(1), Article No. 3.
- Kiverstein, J. D., & Rietveld, E. (2018). Reconceiving Representation-Hungry Cognition: An Ecological-Enactive Proposal. *Adaptive Behavior*, 26(4), 147–163.
- Lakoff, G., & Johnson, M. (2003). *Metaphors We Live By (2nd ed.)*. Chicago: University of Chicago Press.
- Lin, M., Preston, A., Kharrufa, A., & Kong, Z. (2016). Making L2 Learners' Reasoning Skills Visible: The Potential of Computer Supported Collaborative Learning Environments. *THINK SKILLS CREAT*, 22, 303–322.
- Morris, M. R., Huang, A., Paepcke, A., & Winograd, T. (2006). Cooperative Gestures: Multi-User Gestural Interactions for Co-located Groupware. In *Proceedings of CHI'06* (pp. 1201–1210).
- Piper, A. M., Friedman, W., & Hollan, J. D. (2012). Setting the Stage for Embodied Activity: Scientific Discussion around a MultiTouch Tabletop Display. *IJLT*, 7(1), 58–78.
- Piper, A. M., & Hollan, J. D. (2009). Tabletop Displays for Small Group Study: Affordances of Paper and Digital Materials. In *Proceedings of CHI'09* (pp. 1227–1236).
- Shaer, O., Strait, M., Valdes, C., Feng, T., Lintz, M., & Wang, H. (2011). Enhancing Genomic Learning Through Tabletop Interaction. In *Proceedings of CHI'11* (pp. 2817–2826).
- Stofer, K. A. (2016). When a Picture Isn't Worth 1000 Words: Learners Struggle to Find Meaning in Data Visualizations. *Journal of Geoscience Education*, 64(3), 231–241.
- Vygotsky, L. S. (1978). *Mind in Society: Development of Higher Psychological Processes*. Harvard University.
- Wilson, A. D., & Golonka, S. (2013). Embodied Cognition is Not What You Think It Is. *Frontiers in Psychology*, 4, Article No. 58.

## Acknowledgments

This work is partially supported by National Science Foundation Grant Awards #DRL-1612485 and #IIS-1552598. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect these agencies' views. The authors also thank the Florida Museum of Natural History for allowing us to recruit participants from their visitors. This work was conducted while author Jeremy Alexandre was a summer intern at the University of Florida.

# “I Cannot Explain Why I Like This Shape Better Than That Shape”: Intercorporeality in Collaborative Learning

Jacob Davidsen and Thomas Ryberg  
jdavidsen@hum.aau.dk, ryberg@hum.aau.dk  
Aalborg University – eLearning Lab

**Abstract:** While the exchange of utterances between collaborators have been viewed as the primary vehicle for meaning-making in CSCL, we have been encouraged by the CSCL19 theme to expand this understanding to include intercorporeality as way of foregrounding bodies in collaborative learning. We find that collaboration, meaning-making and problem solving is also occurring even when collaborators are lacking words for expressing their thoughts in talk. Thus, we promote the idea of viewing CSCL as a field concerned with intercorporeality in meaning-making practices. Through a detailed analysis of a group of architectural engineering students, we unpack how they prepare for a critique session by building upon each other’s utterances, gestures and manipulations of objects in their design studio thus displaying intercorporeal memory and understanding. Based on the analysis we discuss three themes related to intercorporeality in CSCL: Ecologies of technologies, Historical development of the bodily-material resources and Subtle manifestations of intercorporeal understandings.

## Introduction

In recent years, we have witnessed a re-newed interest in embodied aspects of interaction in CSCL (Davidsen & Ryberg, 2017; Flood, Neff, & Abrahamson, 2015). To some extent this development has been driven by the advent of technologies for supporting collaborative learning such as interactive tabletops (Higgins, Mercier, Burd, & Hatch, 2011), but also from theoretical and methodological orientations towards embodied interaction analysis (Streeck, Goodwin, & LeBaron, 2011). To some degree, the theoretical pillars of CSCL have included and mentioned multimodal and embodied elements of interaction (e.g. Teasley & Roschelle, 1993), but the analysis of the role of the body, the environment and the relations between bodies and environment have been backgrounded in comparison to talk. Encouraged by the CSCL19 theme of 4E learning (Embodied, Enactive, Extended, and Embedded), we add to the re-newed interest in this more ‘holistic’ research agenda of CSCL by presenting a study on the intercorporeality of collaborative learning in ‘architecture and design’ education. Based in our study we suggest that intercorporeality offers a fresh methodological and theoretical stance on collaborative learning by zooming in on how bodies affect each other in collaboration. Thus, this is an attempt to foreground bodies in collaborative learning widening the talk-based focus in CSCL to include and emphasize intercorporeal memory and understandings.

In this paper, we study a ‘design studio’ featuring collaborative activities amongst students in the ‘architecture and design’ program at Aalborg University. The setting and collaborative activities are highly multimodal and composed of both talk, bodies and materials. While many of the activities in the studio are supported by digital technologies, and other highly specialized technologies, we want to broaden the perspective of CSCL by focusing on the material surroundings, bodies, talk, digital and analogue technologies supporting the collaborative activities. The group of six students are in midst of preparing design concepts for an upcoming critique seminar with fellow student groups and teachers. In doing so, they are working together in different collaborative constellations – as a whole group and in pairs depending on the type of activity and work tasks they face. To get a better understanding of how the students are collaboratively preparing for a critique seminar, we study their intercorporeal work in the studio – integrating the Embodied, Enactive, Extended and Embedded aspects into one analysis. Further, we suggest that intercorporeality can function as a theory highlighting the intimate and ‘inter-kinesthetic’ relations between bodies in collaborative learning.

In the paper we are focusing intercorporeality and in particular on how collaboration, meaning-making and problem solving is possible even when the participants lack words for expressing their ideas and arguments. This includes showing how bodies, technologies, objects, gestures, touches, and other bodily-material resources are contributing to the formulation and understanding of ideas and design concepts in creative collaborative design processes between students.

## From talk to embodiment to intercorporeality in CSCL



Since the early years of CSCL, researchers have studied the subtle processes constituting collaborative learning with technologies in face-to-face settings and online learning environments. Much work have focused on talk (spoken or written) between the collaborators, but there is also a growing body of studies integrating embodied elements in the analysis of collaboration. Basically, we see a development in the focus of CSCL studies from talk to embodiment, but our goal with this paper is to add an additional perspective to these established analytical foci by introducing the idea of intercorporeality in collaborative learning as a way of focusing on the intimate, ‘inter-kinesthetic’, and affective bodily aspects of collaboration.

As CSCL is concerned with collaborative activities in both online and physical environments, there is a range of different methodological and theoretical stances available for analyzing and designing for collaborative learning (Wise & Schwarz, 2017). On a general level CSCL is divided between three types of studies: experimental, design and naturalistic (Stahl, Koschmann, & Suthers, 2006). Across the types of studies and characteristics of the learning environments (e.g. face-to-face or online), there is a strong orientation towards analyzing talk. Stahl (2006, p. 6 italics in original) suggested that “*meaning is created across the utterances of different people*”, emphasizing the interest in talk and text in CSCL. However, as argued by Flood, Neff and Abrahamson (2015) there is a need for CSCL to develop ways of “representing and cataloguing choreographies of embodied interaction” (2015, p. 96) in order to better design for collaborative embodied interaction in CSCL environments. One example of such a catalogue of embodied interaction in CSCL is found in Davidsen and Ryberg (2017), who explored children’s collaborative activities in front of a shared touch-screen. The authors suggested that bodily-material resources and in particular hands are used as resources for ‘communicating and illustrating’, as ‘cognitive auxiliary tools’ and as a way ‘shepherding’ each other.

Beyond the field of CSCL there is an extended and growing body of research focusing on embodiment and how the elements in the environment are used by the participants. While these studies are not addressing collaborative learning per se, they are focusing on the ways people accomplish tasks together in various settings and how they competently master a practice. For example, in a series of studies in the past 40 years Goodwin (2017) has shown how people build upon each other’s embodied interactions through different media and resources at archeological sites, court rooms, oceans vessels, etc. In addition, a recent collection of papers (Meyer, Streeck, & Jordan, 2017a) address the notion of intercorporeality in human interaction in diverse practices (e.g. families, self-defense training and auto repair workshops). The paradigm of intercorporeality is first of all rejecting a transmission view of interaction, but it is also opening up the field of embodied interaction by fleshing out that bodies are not alone – they are within an “inter-kinesthetic field” (Behnke, 2008). Meyer, Streeck and Scott (2017b) argue that intercorporeality is “not only the presences, movements, and micromovements of other (inanimate or animate) bodies in my peripersonal space, but also the sedimented traces of such presences and movements in the architectures and the artifacts around my body” (2017b, p. xvii) – what Fuchs (2017) refer to a intercorporeal memory. Meyer et al. (2017) further critique the use of established concepts (also frequently used in the CSCL literature) as ‘coordination’, ‘alignment’, ‘intersubjectivity’ and ‘joint attention’ as they are not doing justice to the synchronous, simultaneous, improvisational and bodily actions of the participants. For instance, in an analysis of ‘Intercorporeality at the motor block’, Alkemeyer, Brümmer and Pille (2017) find that the character and nature of movements in the situation fosters communication and understanding among the members of the practice. The authors suggest that “These respective movements can be understood as a collective way of thinking, reflecting and problem-solving by means of gestural demonstration and experimentation” (Alkemeyer et al., 2017, p. 227).

The paradigm of intercorporeality seems to offer a fresh perspective on collaborative learning mediated by technology. With intercorporeality the lens is widened in CSCL and it is a call to go beyond talk and embodiment to start focusing on the intertwined nature of bodies in collaborative activities. Integrating intercorporeality necessarily also requires a broader view on what can support collaborative learning in CSCL environments. Instead of solely studying interaction with one technology or application, it is necessary to follow the methodological and theoretical guidance from embodied interaction analysis and intercorporeality to study talk, bodies and material surroundings as one complex analytic unit.

## **Context of the study and data collection**

In this paper, we examine a group of ‘architecture and design’ students in their 5<sup>th</sup> semester at Aalborg University. The overall pedagogical idea of Aalborg University is based on Problem Based Learning (PBL). PBL is often considered a pedagogical approach supporting students in obtaining transversal competences such as collaboration, communication, critical thinking and problem-solving skills (Du, Emmersen, Toft, & Sun, 2013; Guerra, 2017). Each semester students will work on a group project supervised by a researcher. Besides receiving feedback and critique from the supervisor the group will also present, discuss and evaluate their project in formal critique sessions. The project is finally evaluated in a group exam where the group will

present, discuss and reflect upon their work together with the supervisor and an external examiner who will ultimately assess and grade the project.

Over a period of six weeks, we video observed three groups working in their open learning environment. We zoom in on one of the groups. The group was allocated a studio in a large open learning environment together with 13 other groups at the semester. Each group decorated and furnished the studio for their specific needs and the group we study in this paper used several notice boards and tables to support their collaborative activities (Figure 1). Each of the notice boards had different purposes, e.g. a shared calendar, a to-do-list, printed photos for inspiration, drawings, etc. On the center table, the students gathered for collective activities, but also to work in pairs or as individuals. On the table in the center, the students had their laptops, different drawings and styrofoam models. Next to the center table, a smaller table was placed for individual work. Over time the room would clutter with all sorts of materials in the studio, telling a unique story of the process of the individual group.



Figure 1. Photo of the group members at the center table.

In the studio, we positioned 4 fixed video cameras and one of the students wore a chest-mounted GoPro on several occasions. As the collected data is extensive (+ 150 hours for this group alone), we have decided to limit our analysis to a period of two days where the group is preparing for an architectural critique session. Specifically, we analyze a 33 seconds long sequence, which serve the purpose of illustrating how they work together in the context of their design studio preparing for a presentation at the critique seminar. Previous studies have analyzed the pedagogical nature of critique sessions in architectural education (Lymer, 2009; Lymer, Ivarsson, & Lindwall, 2009), but in this paper we focus on the work the students do to prepare for the critique session. We find the preparation work particularly interesting for studying the intercorporeality of collaborative learning, as the students have to develop design concepts together for a shared presentation. In addition, the students are working under a certain time pressure, as there is fixed deadline. The preparation work requires working towards a somewhat shared idea of the design concept through collaborative activities managed independently by the students. One day prior to the critique session, the group is narrowing down the number of design concepts they will present. To scaffold this activity the student group split into three pairs – each pair developing a concept to present to the other group members. In this way they are simulating or rehearsing for the actual critique seminar. In the analysis, we zoom in 33 seconds where Heidi and Sine have just presented their new idea to the other group members, and now the others start critiquing their design concept.

### Contextualizing the sequence

Before the analysis of the interaction, we briefly present the overall structure of the sequence followed by a transcript (Figure 3). The transcript builds on conventions from Conversation Analysis (Jefferson, 2004) and in particular the work of Goodwin (2017) in relation to the use of frame grabs in the transcript. Thus, we have included frame grabs from the video footage in the transcript to make the intercorporeal and simultaneous aspects of the collaborative activity ‘somewhat’ visible – each photo is numbered (P.number) for reference in the analysis.

Heidi and Sine just finished presenting their design concept, and now the rest of the group members start critiquing the design concept. The phase of critiquing the design is both oriented towards highlighting ‘flaws’ and collectively developing better ideas.

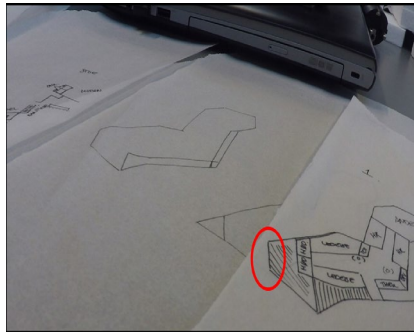
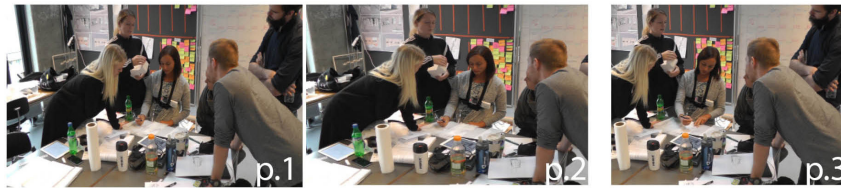


Figure 2. Shape of the building without the triangular tip (marked with the red circle) on the left, the original shape of the building in styrofoam on the right.

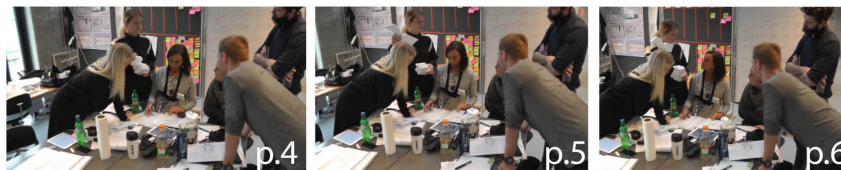
Initially we want to draw attention to the main aspect of the re-design made by Heidi and Sine, who have changed the overall shape of the building by cutting off the triangular tip (see Figure 2 left). This is a quite dramatic change, as the group has been working with the other shape for a couple of days (see Figure 2 right). What follows is an extract of interaction, where the group is discussing the shape of the building and in particular what the triangular tip should accommodate. Ina is evaluating the re-design by picking up the styrofoam model to express her concerns. Ina is also pointing out that they could maintain the shape of the building by extending the window section following the original shape. Much of this discussion is oriented to both the design and the functionality of the triangular tip. While the triangular tip is ‘fulfilling’ the shape of the building it is also a difficult space to use. Ina suggests incorporating a staircase or a transparent elevator shaft in the triangular tip, but again the shape is obstructing her idea. Sven and Sine joke about a transparent elevator with a triangular shape. Now, Ina returns to her defense of the original concept.

### Transcript of the extract

1 I so you like (.) cause I do really like that you like complete the shape in it



2 I cannot explain why I like this shape better than that shape



3 it is just cause I believe it completes the concept

4 H yes

5 I which I believe is very strong right now

6 S it could be like a raised platform over there



7 I yah it could be [something wild growing in there]

8 S [it (.) there could also be something] wild was growing

9 H but that could you also do  
 10 I or a tree  
 11 H yes that that could damn well be tree  
 12 S you just need to say green to Heidi  
 13 All ((laughing))  
 14 M it was actually (.) it is the one you have there with Sines,idea (.) that one with just like a round one

15 SV yes  
 16 M so if you just have that as a square (.) square triangle (.) it could very well be good

17 H I actually think it could be very cool

Figure 3. Transcript of the sequence.

### Analysis: “I cannot explain why I like this shape better than that shape”

While discussing Heidi and Sine’s design concept they use a range of different materials in the environment (Embedded and Extended); a styrofoam model consisting of several separate layers, existing drawings on the tables, noticeboards, an Ipad, and pencils for adjusting drawings or making new ones. The ecology of resources supporting their collaborative activity is diverse and continually configured and arranged to the specific needs of the activity. The combination of materials, bodies and talk facilitate their preparation for the formal critique session. While talk is of course crucial for meaning making, it is also clear, as we shall show, that intercorporeality establishes dimensions of sociality and collective knowing not present in their talk.

In the beginning of the extract (line 1), Ina is stressing that she would like the shape ‘to be completed’ saying “so you like (.) cause I really do like that you like complete the shape of it”, while simultaneously touching (almost embracing) the triangular tip of the styrofoam model with her right hand (P.1, 2 and 3). Actually, Ina repeats her touch of the figure twice, which indicates the importance of keeping the triangular tip and further illustrates what area of the building she is thinking of to the rest of the group members. In line 2, she says “I cannot explain why I like this shape better than that shape (.) it is just cause I believe it completes the concept” and then (p. 4) she separates the top layer of the styrofoam model from the rest by twisting it so the two-dimensional layout of the building is facing her fellow group members (p. 5 and 6). Ina is not utilizing any scientific or academic terms in trying to convince the other group members about keeping the triangular tip. Mostly, she is relying on everyday language combined with her bodily manipulation of the styrofoam model in front of the other group members. When she separates the layers and say “which I believe is very strong right now“ (line 5), she is further linking the styrofoam model with the design concept developed by the group. The ‘flat’ two dimensional figure of the building is similar to the one on paper in front of Heidi and Sine, which could indicate that Ina is coupling the two different forms of the models.

The intercorporeal understanding between the students furthermore become visible as the other group members build on Ina’s insistence on keeping the triangular shape of the building. First Sven, points to the drawing in front of Heidi saying “it could be like a raised platform over there” (line 6, P. 7), which is

establishing a connection between the styrofoam model in Ina's hands and the drawing on the paper in front of Heidi. Then Ina adds the idea of something green, saying "yah it could be something wild growing in there" in overlap with Sven saying "it (.) there could also be something wild was growing". Sven is further embodying the idea of a tree by moving his hands up into the air showing the shape of a tree (P. 8-12). Ina replicates a similar movement raising both her hands in the air looking in the direction of Sven (P. 10). The concrete manifestation of the tree shows that their understanding is "based on embodied intersubjectivity" (Fuchs, 2017, p. 17). Heidi accepts their suggestion saying "but that you could also do", but then Ina makes the suggestion more concrete saying "or a tree", which is followed by Heidi saying "yes that that could damn well be a tree". Sine then says "you just need to say green to Heidi" indicating that Heidi is favoring green/wild elements in their design. In response to this Heidi, Sine and Sven starts laughing.

Then, in the final part of the sequence after Heidi, Sine and Sven laughed (line 13), Mette builds on the interactional work by Ina, Heidi and Sven by referring to a previous drawing (made by Sine) hanging on the notice board opposite to where the group is located. Mette says "it was actually (.) it is the one you have there with Sines idea (.) that one with just like a round one" pointing to the area of the notice board (P. 13) (later – not part of this sequence - Ina is including Sine's drawing again, but she walks to the noticeboard and points to the drawing to integrate it into their discussion). Thus, the collective spatio-temporal history of the group is made present and relevant for their current work on the functionality of the triangular tip of the building. The original idea drawn by Sine is a cylinder shape going through the floors of a building (P. 13), which is now transformed into an idea illustrating a possible functionality of the triangular tip. Mette elaborates "so if you just have that as a square (.) square triangle (.) it could very well be good" while she is making imaginary cuts and environmentally coupled gestures (Goodwin, 2007) on the styrofoam model with her left hand. The cuts are detailing the triangular tip on the styrofoam model and the gestures are showing the idea of adding something green going through the floors of the building. These imaginary cuts and gestures are repeated twice through P. 18-22 as a way of communicating and illustrating the connection between the drawing on the notice board and the idea of adding something green (a tree) in the triangular tip.

Throughout the 33 seconds of interaction the 4Es are in play in several ways: the group is building on the immediate presence of each other's bodies (Embodied) and the history Embedded in the material surroundings (Extended and Enactive). In addition, it is visible that the students repeat their movements as a way of arguing, problem solving and designing together. They display what could be referred to as an intercorporeal form of knowing (Hindmarsh & Pilnick, 2007). As noted by Jornet and Steier (2015) such bodily engagements constitute the 'infrastructure' supporting collaborative activities over time. Thereby, talk, bodies and material surrounding mutually provide the ground for their collective designing and thinking – from the initial idea of something green to something wild and then finally a tree. Further, it is also evident that the students refer to and make use of an ecology of technologies – the infrastructure supporting their collaborative activity is manifest in many different mediational resources.

## Discussion

In the following sections, we broaden the analytical focus by discussing three themes, which could have an impact on the CSCL research agenda: Ecologies of technologies, Historical development of the bodily-material resources and Subtle manifestations of intercorporeal understandings.

### Ecologies of technologies

As argued by Suthers (2006) CSCL is a field consisting of a learning element and a technology element. Based in our analysis, it seems that CSCL should not only be concerned with digital technologies in collaborative learning. In the sequence, the students are using many different technologies and materials to support their collaborative activities. In line with the 4E agenda, it seems relevant that CSCL studies in "natural settings" (Stahl et al., 2006) should open and widen the analytical focus to more than just one computer application. This include studying and designing technologies embedded in a particular environment, but also to accept that supporting collaborative learning is complex in nature, i.e. the computer application is not the only variable influencing the processes of collaborative learning.

### Historical development of the bodily-material resources

A crucial moment in the students' collaborative activity is happening when Mette is referring to and including a drawing made by Sine previously. To paraphrase Goodwin (2017) they are making use of materials made available from the predecessors by re-instantiating the drawing in their current activity. This is in line with Meyer et al. (2017a, p. xvii) arguing to incorporate the "sedimented traces of such presences and movements in the architectures and the artifacts around my body" in analyzing intercorporeality. In these face-to-face settings



it seems highly relevant to study more than the “*meaning (...) created across the utterances of different people*”. (Stahl, 2006, p. 6 italics in original). Following the line of work informed by embodied interaction analysis (Streeck et al., 2011) and intercorporeality (Meyer et al., 2017a), we find it necessary to incorporate a more historical and developmental view of the bodily-material resources that collaborators integrate and build upon in their collaborative activities. In the sequence with the architecture and design students, we find that students are making certain bodily-material resources relevant in their interaction for supporting their collaborative work, e.g. Ina’s repeated touch of the triangular tip, Ina and Sven embodying the shape of a tree and Mette’s reference to the spatio-temporal resources in the studio. This indicates intercorporeal understanding and memory in the collaborative activities.

### Subtle manifestations of intercorporeal understandings

The subtle manipulation of the styrofoam model indicates an intercorporeal understanding among the students and it seems the imaginary cuts and gestures are arguments in their own right in this collaborative activity. While, the video is allowing access to the subtle bodily manifestations of intercorporeal understanding, we are still limited in our way of representing these details in a paper format. As argued by Flood et al. (2015) there is a need for CSCL to develop ways of “representing and cataloguing choreographies of embodied interaction” (2015, p. 96). In most CSCL studies the product of representation is manifested in a transcript (sometimes including frame grabs from the videos). In this paper, we also rely on a transformation from video to text, which is hardly showing the repeating gestures and touches on the styrofoam model. Thus, CSCL need to further develop ways of making the subtle manifestations of intercorporeal understandings visible to better understand the nature of collaborative learning and to design better environments for CSCL in the future.

### Conclusion

With this paper we seek to broaden the methodological and theoretical cartography of CSCL by introducing the concept of intercorporeality in collaborative learning. With intercorporeality we explore the relations between the Embodied, Enactive, Extended, and Embedded aspects of CSCL. Further, we are able to zoom in on the relations of bodies and how they affect each other – studying their ‘inter-kinesthetic fields’ (Behnke, 2008) in relation to collaborative activities. In the analysis, we have shown that the students in preparing for a collective presentation of design concepts are embodied and embedded in their local environment using materials to foster intercorporeal understandings. Furthermore, that the students are making sense through talk, bodies and the material surroundings which show signs of their intercorporeal understanding. Most importantly, we made visible that the students understand each other through their gestures and touches on the shared material (e.g. the styrofoam model). They are able to collaborate and problem solve even when they are lacking words for expressing their ideas. This aligns with the findings from Hahn and Jordan (2017), who suggested that objects could foster and extend bodily meaning-making processes.

Historically, CSCL has been concerned with “meaning and the practices of meaning making in the context of joint activity” (Koschmann, 2002, p. 20) manifested in primarily language. In the future, CSCL could be oriented towards unpacking the intercorporeal dimensions of collaboration, meaning-making and problem solving. We believe that this paper is a modest contribution in expanding the theoretical, methodological and analytical unit of CSCL to include and emphasize how collaborators make sense and solve problems in an intercorporeal modus. In future studies, it would be of great interest to revisit some of the established theories in CSCL (e.g. Group Cognition (Stahl, 2006) and Knowledge Building (Scardamalia & Bereiter, 1994) through the lens offered by intercorporeality.

### References

- Alkemeyer, T., Brümmer, K., & Pille, T. (2017). Intercorporeality at the motor block - On the importance of a practical sense for social cooperation and coordination. In C. Meyer, J. Streeck, & J. S. Jordan (Eds.), *Intercorporeality: emerging socialities in interaction* (pp. 203–235). New York, NY: Oxford University Press.
- Davidson, J., & Ryberg, T. (2017). “This is the size of one meter”: Children’s bodily-material collaboration. *International Journal of Computer-Supported Collaborative Learning*, 12(1), 65–90.
- Behnke, E. A. (2008). Interkinaesthetic Affectivity: A Phenomenological Approach. *Continental Philosophy Review*, 41(2), 143–161.
- Du, X., Emmersen, J., Toft, E., & Sun, B. (2013). PBL and critical thinking disposition in Chinese medical students – A randomized cross-sectional study. *Journal of Problem Based Learning in Higher Education*, 1(1), 72–83. <https://doi.org/10.5278/ojs.jpblhe.v1i1.275>

- Flood, V. J., Neff, M., & Abrahamson, D. (2015). Boundary Interactions: Resolving Interdisciplinary Collaboration Challenges Using Digitized Embodied Performances. In O. Lindwall, P. Häkkinen, T. Koschmann, P. Tchounikine, & S. Ludvigsen (Eds.), *Exploring the Material Conditions of Learning: Opportunities and Challenges for CSCL* (Vol. Vol 1, pp. 94–100). Gothenburg: The international Society of the Learning Sciences.
- Fuchs, T. (2017). Intercorporeality and interaffectivity. In C. Meyer, J. Streeck, & J. S. Jordan (Eds.), *Intercorporeality: emerging socialities in interaction* (pp. 3–23). New York, NY: Oxford University Press.
- Goodwin, C. (2007). Environmentally coupled gestures. In D. McNeill, S. D. Duncan, J. Cassell, & E. T. Levy (Eds.), *Gesture and the dynamic dimension of language: essays in honor of David McNeill* (pp. 195–212). Amsterdam ; Philadelphia: J. Benjamins Pub. Co.
- Goodwin, C. (2017). *Co-operative action*. Cambridge University Press.
- Guerra, A. (2017). Engineering grand challenges and the attributes of the global engineer. *Proceedings of the 45th Sefi Annual Conference 2017*, 1222–1235.
- Hahn, T., & Jordan, J. S. (2017). Sensible objects: Intercorporeality and enactive knowing through things. In C. Meyer, J. Streeck, & J. S. Jordan (Eds.), *Intercorporeality: emerging socialities in interaction* (pp. 267–287). New York, NY: Oxford University Press.
- Higgins, S., Mercier, E., Burd, E., & Hatch, A. (2011). Multi-touch tables and the relationship with collaborative classroom pedagogies: A synthetic review. *International Journal of Computer-Supported Collaborative Learning*, 6(4), 515–538. <https://doi.org/10.1007/s11412-011-9131-y>
- Hindmarsh, J., & Pilnick, A. (2007). Knowing Bodies at Work: Embodiment and Ephemeral Teamwork in Anaesthesia. *Organization Studies*, 28(9), 1395–1416. <https://doi.org/10.1177/0170840607068258>
- Jefferson, G. (2004). Glossary of transcript symbols with an introduction. In G. H. Lerner (Ed.), *Conversation Analysis: Studies from the first generation* (Vol. 125, pp. 13–34). Amsterdam ; Philadelphia: John Benjamins Publishing Company.
- Jornet, A., & Steier, R. (2015). The Matter of Space: Bodily Performances and the Emergence of Boundary Objects During Multidisciplinary Design Meetings. *Mind, Culture, and Activity*, 22(2), 129–151.
- Koschmann, T. (2002). Dewey's contribution to the foundations of CSCL research. In *Proceedings of the Conference on Computer Support for Collaborative Learning: Foundations for a CSCL Community* (pp. 17–22). International Society of the Learning Sciences.
- Lymer, G. (2009). Demonstrating Professional Vision: The Work of Critique in Architectural Education. *Mind, Culture, and Activity*, 16(2), 145–171. <https://doi.org/10.1080/10749030802590580>
- Lymer, G., Ivarsson, J., & Lindwall, O. (2009). Contrasting the use of tools for presentation and critique: Some cases from architectural education. *International Journal of Computer-Supported Collaborative Learning*, 4(4), 423–444. <https://doi.org/10.1007/s11412-009-9073-9>
- Meyer, C., Streeck, J., & Jordan, J. S. (Eds.). (2017a). *Intercorporeality: emerging socialities in interaction*. New York, NY: Oxford University Press.
- Meyer, C., Streeck, J., & Jordan, J. S. (2017b). Introduction. In C. Meyer, J. Streeck, & J. S. Jordan (Eds.), *Intercorporeality: emerging socialities in interaction* (pp. xv–xlix). New York, NY: Oxford University Press.
- Scardamalia, M., & Bereiter, C. (1994). Computer Support for Knowledge-Building Communities. *Journal of the Learning Sciences*, 3(3), 265–283. [https://doi.org/10.1207/s15327809jls0303\\_3](https://doi.org/10.1207/s15327809jls0303_3)
- Stahl, G. (2006). *Group cognition computer support for building collaborative knowledge*. Cambridge, Mass.: MIT Press.
- Stahl, G., Koschmann, T., & Suthers, D. (2006). CSCL: An Historical Perspective Computer-supported collaborative learning. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 409–426). Cambridge handbook of the learning sciences. Cambridge, UK: Cambridge University Press.
- Streeck, J., Goodwin, C., & LeBaron, C. (2011). Embodied interaction in the material world: An introduction. In J. Streeck, C. Goodwin, & C. D. LeBaron (Eds.), *Embodied interaction : language and body in the material world* (pp. 1–26). New York: Cambridge University Press.
- Suthers, D. D. (2006). Technology affordances for intersubjective meaning making: A research agenda for CSCL. *International Journal of Computer-Supported Collaborative Learning*, 1(3), 315–337.
- Teasley, S. D., & Roschelle, J. (1993). Constructing a Joint Problem Space: The Computer as a Tool for Sharing Knowledge. In C. O'Malley (Ed.), *Computer-supported collaborative learning* (pp. 229–258). Springer New York.
- Wise, A. F., & Schwarz, B. B. (2017). Visions of CSCL: eight provocations for the future of the field. *International Journal of Computer-Supported Collaborative Learning*, 12(4), 423–467.

# Effects of the Need for Cognitive Closure and Guidance on Contribution Quality in Wiki-Based Learning

Sven Heimbuch and Daniel Bodemer  
sven.heimbuch@uni-due.de, bodemer@uni-due.de  
University of Duisburg-Essen, Germany

**Abstract:** Controversies based on opposing points of view and contradictory evidence can be fruitful to trigger individual elaboration processes. However, research showed that many wikis are not necessarily suited to make relevant content salient and thus users need further guidance. In an experimental laboratory study on wikis, we investigated two guidance types in conjunction with the Need for Cognitive Closure: (1) visual markers to highlight controversy status (implicit guidance) and (2) a collaboration script that directs users towards discussions (explicit guidance). We analysed the quality of  $N = 724$  wiki contributions, namely final edits to an article and corresponding discussion replies. The results show that both guidance types do neither directly affect the user contributions to the wiki nor the learning outcome. In interaction with the individual Need for Cognitive Closure there was a meaningful effect on the quality of discussion contributions, matching previous research findings on the quantitative learning outcomes.

## Introduction and background

Working and learning productively and efficiently in groups is becoming increasingly important (Miller & Hadwin, 2015). Net-based communication can help by maintaining the benefits of offline communication, such as mutual knowledge sharing (McLoughlin, 2002), while reducing disadvantages, such as social inhibitions (Murray, 2003). However, it is not only possible to use net-based environments for communication, but it can also support users in their individual learning processes and (cooperative) knowledge construction (Cress & Kimmerle, 2008). A well-known and popular environment, which offers the possibility of cooperative knowledge construction, are wiki pages and the corresponding discussion pages. These allow participants to participate in existing discussions, i.e. express themselves or even clean up a topic, initiate new discussions about topics that are important to them, and finally make changes to the article, ideally based on the discussions about an article. Such collaborative article writing is one of the most common group activities (Schlichter, Koch, & Bürger, 1998). In a wiki environment, it is not unusual or unintended for conflicts or constructive controversies to occur (Johnson, Johnson, & Tjosvold, 2000) when the wiki information deviates from one's personal knowledge (Cress & Kimmerle, 2007). This makes it necessary to assimilate or accommodate internally from the user's perspective or externally from the wiki's perspective. This shows that precisely these conflicts can be particularly helpful for the knowledge construction and learning success of the individual (Piaget, 1977). Furthermore, they can also lead to partners being more motivated and using their expertise more effectively (Johnson, Johnson, & Smith, 2000).

The Need for Cognitive Closure (NCC) is an important influence on people's behaviour in dealing with controversies. It is the need that people want to find an obvious solution, although it does not always have to be the right solution (Dijksterhuis, van Knippenberg, Kruglanski, & Schaper, 1996). People with a high NCC strive to achieve a quick solution and show insecure aversive behaviour. However, people with a lower NCC tend to prefer finding the best solution to ambiguity (Schumpe et al., 2017; Webster & Kruglanski, 1997). This difference between high and low NCC persons can also be measured in the speed of decisions, as people with high NCCs perceive ambiguity as unpleasant. Consequently, they tend to make decisions more quickly and use less information and anchor heuristics for judgment formation (Schlink & Walther, 2007). In contrast to that, people with a low NCC enjoy the uncertainty of ambiguous situations. Thus, they tend to postpone the decision and prefer weighing and finding more information (Schlink & Walther, 2007). In wiki-based learning settings, individuals with a low NCC are more likely to search purposefully for additional in-depth information about a topic in an ambiguous situation, whereas high NCC persons will more likely process the simplest information at hand. Although there are close ties between the NCC and inter-individual differences in learning and knowledge construction, there are only few studies in technology-enhanced learning addressing this construct.

With increasing complexity of digital learning environments with, it can become helpful that learners are further supported in dealing with controversial and ambiguous information with the help of supplemental *Cognitive Group Awareness* tools (Bodemer & Buder, 2006). Although minimal tacit guidance for learners has been questioned by others (Kirschner, Sweller, & Clark, 2006), we do not fully agree to such



floccinaucinihilipilification of implicit guidance. Cognitive group awareness tools that are focused on gathering and visualising knowledge-related contextual cues have been successfully implemented as implicit guidance measures to structure collaborative learning processes (Heimbuch & Bodemer, 2017). Another line of wiki-related research has proposed additional measures of explicit guidance to incorporate in wiki-based learning environments to improve the overall quality of knowledge artefacts and for better coordination processes of students. The implementation of collaboration scripts is one possible explicit guidance measure where the activities of writers and editors within a social system are coordinated and optimised. A script is a set of instructions that specifies the group formation, modes of interaction and task management between collaboration partners (Dillenbourg, 2002). Positive effects have been found for scripts with a special focus on article editing and revising that ultimately led to more coherent articles and fewer inaccurate articles (Wichmann & Rummel, 2013). Furthermore, it has also been shown that a certain level of coercion in the implemented collaboration script is recommended to produce content of higher quality (Papadopoulos, Demetriadis, & Weinberger, 2013).

In previous analyses of the underlying study data, we could show that test persons in the experimental study groups (implicit vs. explicit or high vs. low NCC) showed no meaningful differences in the process variables with the log data that we measured (e.g., topic selection, time to contribute, topic reply frequency), as assumed in the hypotheses (Heimbuch & Bodemer, 2018). Likewise, there was no direct effect of the guidance type on the performance in the knowledge test. The in-depth investigation of the interaction of guidance type and NCC confirmed the hypothesised pattern. Evidence suggested that subjects with a low NCC achieved higher test scores with explicit guidance (collaboration script) and accordingly subjects with a relatively high NCC achieved better scores with implicit guidance (controversy awareness highlights). These findings support our claim that implicit guidance provides a quick, non-restrictive way to find a solution, which is preferred by those persons with a high NCC. Explicit guidance can be beneficial to people with a low NCC, as it can support them to find a better solution rather than just the quickest. To complete our previous analyses on the quantitative part, we were left with the question if we can find promising effects on the quality of contributions when we specifically investigate implicit and explicit guidance in the interplay with the Need for Cognitive Closure. Thus, we were interested in two main research questions:

*RQ 1:* What kind of contributions would we find in either the implicit guidance or the explicit guidance wiki group, and how can these be translated into a quality measure?

*RQ 2:* Does the type of guidance directly or indirectly influence the quality of contributions and learning outcomes, and how does the NCC moderate potential effects?

## Method

### Participants and study design

The  $N = 181$  participants were mostly undergraduate students at the University of Duisburg-Essen (Germany) in their first semester. Students were recruited via the university's social media channels and on-site. Their age range spanned from 17 to 33 years ( $M = 20.59$ ,  $SD = 2.59$ ). We have randomly assigned students to one of two experimental learning environments about different forms of energy (e.g., nuclear power, fossil fuels, renewable energy). This was either a wiki with added controversy awareness highlights for implicit guidance or a wiki with a collaboration script for explicit guidance. Thus, this variation of wikis was our main independent variable. Furthermore, we have assessed the Need for Cognitive Closure with the German short scale 16-NCCS (Schlink & Walther, 2007) of each participant and used this as a second factor for analytical purposes. Our main dependent variable was the article and discussion quality of each student's contribution. Moreover, we have also measured numerous process variables through data logging and learning outcomes with knowledge tests.

### Study procedure

We conducted the experiment in an individual setup with up to four participants at the same time, separated by divider panels. Participants performed all the experiment's stages individually in their own wiki instance. After participants were briefed with standardised written instructions on the computer screen and had given consent to participate in the study, they were first asked a few basic socio-demographics as well as interest in and prior knowledge of the study's subject matter (forms of energy). This was followed by a short mandatory introduction to the specific wiki environment. We asked the participants to click through a mock-up environment with "lorem ipsum" texts to familiarize with the general wiki structure. In addition to the general orientation in a wiki, this tutorial phase also served to familiarize with the specific additions we added to the experimental wikis (controversy highlighting vs collaboration script) to ensure that participants have a common ground about their wiki environment's mechanics (Figure 1).



Figure 1. In the floating upper left part of the figure, we illustrate the wiki's talk page table of contents with controversy status highlights for implicit guidance. On the right, we illustrate a collaboration script representation as used for the explicit guidance wiki.

Both groups had the same task of contributing to an initial Wikipedia-like base article about different forms of energy and participating in up to three of the corresponding discussions. Participants received the information that the discussions contain enough arguments and evidence to enrich the original article, since we did not provide any other additional material regarding the subject matter elsewhere. We gave no further instructions on how to start their wiki task (e.g., reading the article or any discussion first) or what kind of reply they should make to a self-selected discussion. This was the experiment's main stage where participants had a loose total time limit of 21 minutes for finishing all article edits and discussion replies. After the time for a contribution phase was up, they the environment automatically prompted them to finish their contributions in the wiki and proceed further. Followed by the wiki contribution stage, we provided them with the questionnaires to determine their individual levels of their Need for Cognitive Closure (16-NCCS). After filling out these questionnaires participants had to answer a multiple-choice test about the study's contents. As an additional manipulation check, we asked participants to sum up briefly in open text fields why they have selected certain discussions to comment on and what led to the final decisions for the resulting article edits. Finally, to gain insights about how participants evaluate the additions we made to the wikis we asked them to fill out the User Experience Questionnaire (UEQ) by Laugwitz, Held, & Schrepp (2008).

### Wiki contribution coding

In total, we had to analyse  $N = 181$  article edits and  $N = 543$  replies to the article's discussions. We decided to start our content categorisation deductively with previously discussed categorisations of user types in wikis who participate in the co-evolution of knowledge (Cress & Kimmerle, 2008). The first two categories for edits to the article that we have derived were *Accommodators* and *Assimilators*. The former is mainly characterised as a user whose contributions are mostly restructuring and synthesizing tasks, whereas the latter in the simplest form is purely adding new content to the wiki and not caring much about purposeful integration (Majchrzak, Wagner, & Yates, 2006). With further inductively derived categorisations for article edits, we added four additional categories: *Reformulator*, *Reformulator-shortener*, *Shortener* and *No edit*. Subsequently, we used these final six article editing categorisations for our quality assignments (Table 1). A second trained coder was asked to categorise and rate a random sample of 12 articles and resulted in a  $K = .79$ , which is an adequate level of concordance (Mayring, 2015).

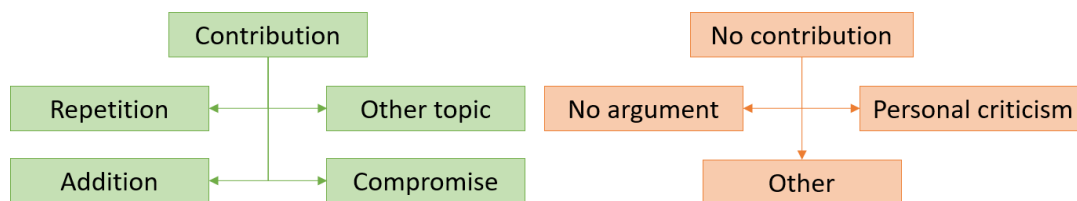
As a second step of our quality analyses, we had to work through the discussion replies and assign adequate categories to these contributions. We decided to work fully inductive with the material in multiple iterations, because we had to expect a large diversity in reply content and quality. Finally, we ended up with seven discussion reply categories: *No statement*, *Repetition*, *Addition*, *Compromise*, *Other (related) topic*, *Personal criticism*, *Other*. Furthermore, we arranged the discussion reply categories in a flat hierarchy. Therefore, we created two higher-level categories for participants' replies, namely test persons who contributed

to the course of the discussion with their comments and those who replied to discussion threads without advancing the discussion by any means (Figure 2).

**Table 1: Assignment of categories to article quality ranks (with justifications)**

| Quality rank | Designation                | Description   |
|--------------|----------------------------|---|
| 6            | Much accommodation         | - Major revisions<br>- More article structuring                     |
| 5            | Little accommodation       | - Minor revisions<br>- Minor article structuring                    |
| 4            | Much assimilation          | - Added more content  |
| 3            | Little assimilation        | - Added some content  |
| 2            | Reformulated and shortened | - Combined reformulation and shortening of sections                 |
| 1            | Reformulated or shortened  | - Minor reformulation of sections<br>- Minor shortening of sections |
| 0            | No edit                    | - No visible text changes   |

Note. Orange = lower quality, yellow = medium quality, green = higher quality



**Figure 2.** Schematic representation of the hierarchy of the categories into two main categories and a total of seven subcategories.

In contrast to the quality assignments for the final article edits, we assigned quality ranks to discussion patterns. The order in which the categories appeared in the three discussions that each participant had to contribute was irrelevant. In order to obtain a quality score for a participant over the course of three discussion contributions, in a next step, we further examined the patterns occurring in the discussion. As can be seen in Table 2, we assigned the highest quality level (rank 9) to a person who brought three new arguments as *Additions* into the discussions. A second trained coder was asked to categorise and rate a random sample of 15 discussion contributions and resulted in a  $K = .74$ , which is again an acceptable level of concordance.

**Table 2: Assignment of categories to quality ranks for discussion triads**

| Quality rank | Discussion triad (order irrelevant)   |
|--------------|---|
| 9            | - Addition – Addition – Addition / Repetition / Compromise  |
| 8            | - Addition – Addition – Other (related) topic / No contribution   |
| 7            | - Compromise – Compromise – Compromise / Addition / Repetition  |
| 6            | - Compromise – Compromise – Other (related) topic / No contribution   |
| 5            | - Addition – Compromise – Repetition / No contribution  |
| 4            | - Addition – Repetition – Other (related) topic / No contribution   |
| 3            | - Repetition – Repetition – Repetition / Addition / Compromise  |
| 2            | - Repetition – Repetition – Other (related) topic / No contribution<br>- Other (related) topic - Other (related) topic – Addition / Compromise / Repetition |
| 1            | - Repetition – Other (related) topic – Compromise / No contribution   |
| 0            | - No contribution – No contribution – Addition / Repetition   |

Note. Orange = lower quality, yellow = medium quality, green = higher quality

## Path analysis

Since this study had a serial process in the action steps of the test persons (three separate iterations of discussing and editing), a path analysis in form of a serial mediation offered itself for the investigation of the effects (Hayes, 2018; Hayes, Montoya, & Rockwood, 2017). Thus, we investigated the path effects of the quality assignments as serial mediators in a row on the effect of the independent variable (type of additional

wiki guidance) on the knowledge test scores as learning outcome, which has been previously analysed in a different context (Heimbuch & Bodemer, 2018).

The score in the knowledge test achieved by the test persons served as the dependent variable. The knowledge test consisted of 18 multiple-choice questions (up to three distractors, at least one attractor) relating to the subject matter of energy sources as presented in the discussions and the wiki article. Furthermore, in order to investigate potential influences of the Need for Cognitive Closure, we performed a moderated serial mediation with discussion quality as first mediator and article edit quality as second mediator. We have chosen this mediator order after investigating the log files and confirming that all participants started with a discussion before an article edit was performed.

## Results and discussion

In order to provide a brief insight into the relationship between the quality allocations of the systems, Table 3 presents an excerpt of the results of both category systems side by side. Two category systems side by side. Even at first glance, a wide variety of combinations of quality ranks can be determined based on the colour coding of the quality ranks. For example, there are similarities between the two quality assignments, as in the case of test person 97, who contributed on a high-quality level in both cases (both green). A match of the quality assignments can also be found with person 9 (both orange) or 96 (both yellow). But there are also differences, as in the case of person 99, whose contributions to the discussion were classified as high quality (green) but the discussion replies were classified as low quality (orange). Overall, in the whole coded dataset we see a relatively even distribution of ranks and user type classifications across both types of guidance.

Table 3. Sample section of the quality assignments in both category systems side by side

| Person | Guidance type | Rank of article quality |        | Rank of discussion quality |        |
|--------|---------------|-------------------------|--------|----------------------------|--------|
| 1      | Implicit      | 3                       | Yellow | 3                          | Orange |
| 2      | Implicit      | 3                       | Yellow | 5                          | Yellow |
| 3      | Implicit      | 3                       | Yellow | 9                          | Green  |
| 4      | Implicit      | 4                       | Green  | 9                          | Green  |
| 5      | Implicit      | 1                       | Orange | 9                          | Green  |
| 6      | Implicit      | 4                       | Green  | 7                          | Green  |
| 7      | Implicit      | 4                       | Green  | 3                          | Orange |
| 8      | Implicit      | 3                       | Yellow | 9                          | Green  |
| 9      | Implicit      | 2                       | Orange | 2                          | Orange |
| ...    |               |                         |        |                            |        |
| 92     | Explicit      | 3                       | Yellow | 5                          | Yellow |
| 93     | Explicit      | 5                       | Green  | 9                          | Green  |
| 94     | Explicit      | 4                       | Green  | 2                          | Orange |
| 95     | Explicit      | 3                       | Yellow | 4                          | Yellow |
| 96     | Explicit      | 3                       | Yellow | 5                          | Yellow |
| 97     | Explicit      | 4                       | Green  | 7                          | Green  |
| 98     | Explicit      | 4                       | Green  | 5                          | Yellow |
| 99     | Explicit      | 4                       | Green  | 3                          | Orange |
| ...    |               |                         |        |                            |        |

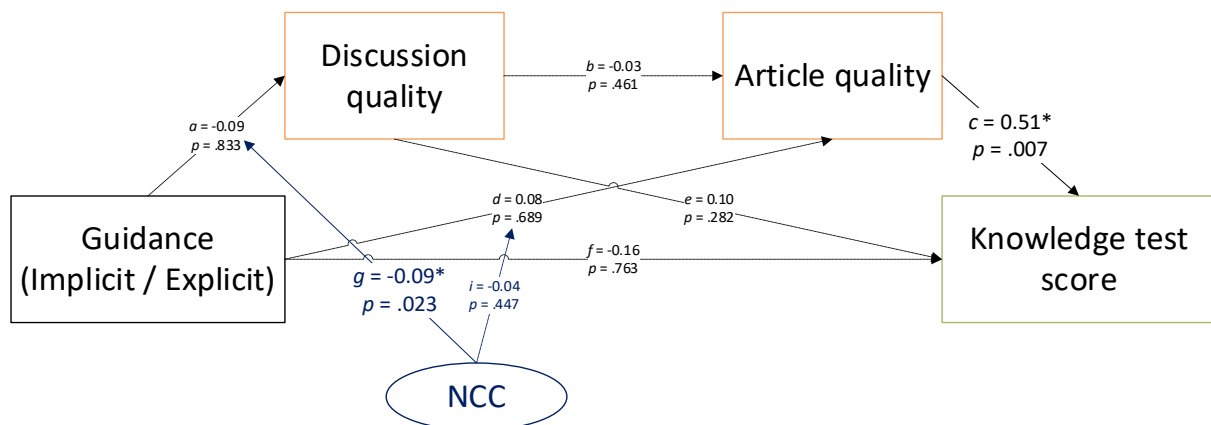
The assigned user categories showed us that the most frequently assigned category was that of *assimilators*, which translated into a high frequency of medium quality contributions to the final article. We assume that social inhibitions could also exist in such a web-based study setting and that the test person must get used to the situation before he or she is "ready" for collaborative knowledge construction (Cole, 2009; Kump, Moskaliuk, Dennerlein, & Ley, 2013). This may have had an influence on the procedure of the test persons to the extent that they did not want to make any more far-reaching changes to the article. As outlined in Table 3, there are no major quantitative differences between the groups with implicit or explicit support in terms of corresponding quality distribution. This is further supported by the results of the path analysis in the following subsection. The path effect of the guidance type on the article editing quality was very small and far from any acceptable level of significance (cf. Figure 2, path *d*), which means that the guidance type had no meaningful direct effect on the final article quality. It should be noted, of course, that since these results were based on a qualitative content analysis, they depend very much on the underlying data. However, since similar categories

have already been found in other studies (e.g., Majchrzak et al., 2006), this indicates that there is a certain generalisability.

When we further investigated the discussion categorisations, we saw the highest frequencies of discussion replies as *additions* and *repetitions*, corresponding to higher and medium quality ranks in our assignments. On the one hand, a high number of persons in the *repetition* category could be explained by superficial processing that would require a rather small amount of cognitive effort (Vertzberger, 1990) to reiterate the arguments of the other participants and thus to write one's own statement. Therefore, if the subjects had not been sufficiently motivated (Cole, 2009) or involved to engage in high cognitive (Newman, Webb, & Cochrane, 1995), they might have relied on behaviour that required only low cognitive effort. On the other hand, the high number of persons in the category *additions* could be explained by the fact that at least some of our test persons were interested and motivated to engage in deeper, more complex behaviour and could therefore ultimately participate in higher quality discussions (Newman, Webb, & Cochrane, 1995). In addition, they may have already had previous knowledge or a firm opinion on the subject in question and have written their reply accordingly.

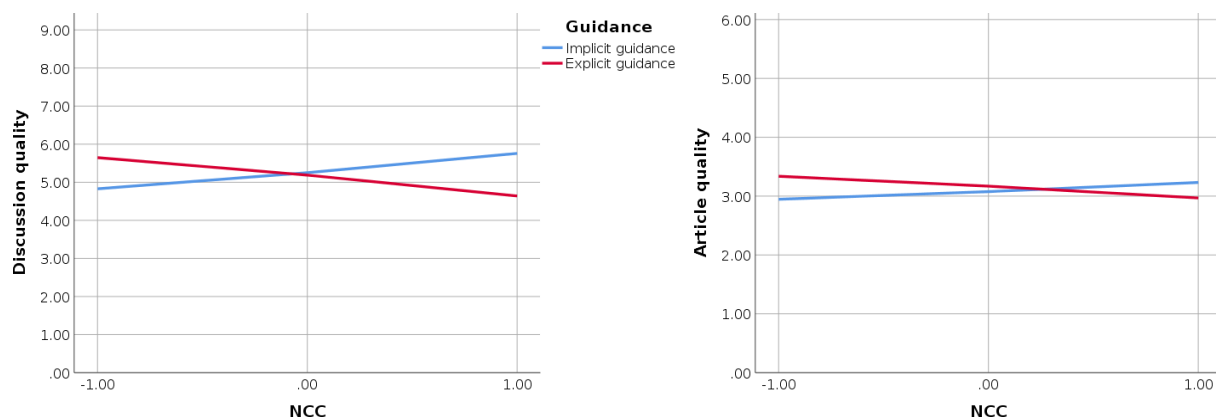
### Path analysis

For analysing the potential moderating influences of the NCC on contribution quality and learning outcomes, we have calculated a moderated serial mediation with both discussion quality and article editing quality as mediating variables. These serial mediations were calculated using Model 85 in PROCESS (v3.1) for SPSS (Hayes, 2018). Figure 3 shows the results of the paths in this model where we show that only one of the unmoderated paths has become significant, namely the path from the article editing quality to the knowledge test score with a regression weight of  $c = .51, p = .007$ . From this it can be concluded that there is a meaningful regression of the article editing quality on the knowledge test score, translating into a test score improvement of approximately 0.5 for each raise in the quality rank. Furthermore, one of the paths moderated by the NCC in this analysis was identified to be of further interest. We found a moderating effect of the NCC on the connection between the guidance type and discussion quality,  $g = -.09, p = .023$ .



**Figure 3.** Representation of the path analysis as moderated serial mediation.

In addition to the above model, we further investigated the simple slopes of the moderations on the paths *a* and *d* (Figure 4). Implicit guidance together with a high NCC and explicit guidance in conjunction with a low NCC can lead to a higher quality of discussion contributions. This matches to the results that were previously presented in the quantitative analyses of the process variables and the learning outcomes (Heimbuch & Bodemer, 2018). This finding can be explained by theories on the NCC and backed by empirical evidence, where persons with a higher NCC want to find rather quick solutions and tend to prefer simpler signals and cue heuristics, whereas low NCC persons enjoy to find better answers and solutions through information seeking and more elaborate discussions (Schlink & Walther, 2007; Webster & Kruglanski, 1994, 1997). The same pattern can also be seen for the article editing quality, although to a much weaker extent. It is reasonable to assume that the effect on this article editing is less obvious, because the NCC refers to dealing with controversies and decision-making situations (Dijksterhuis et al., 1996), which were not necessarily present during the article editing phase. That could be a valid reason why the NCC level had no meaningful moderating effect on the article editing slope.



**Figure 4.** Representations of the moderator effects of the NCC on both quality rank assignments. -1 = low NCC; 1 = high NCC.

## Conclusions

Qualitative content analysis itself was often criticized because of the restriction to rather fixed categories and that this would take the focus from the wholeness of the texts and direct it to paraphrases in the text (Ramsenthaler, 2013). This criticism, however, is debatable and has been refuted by others, since the categories would be re-examined and re-evaluated repeatedly (Mayring, 2014). In our analysis, we categorised entire articles and discussion contributions, which in turn ensured a more holistic view of the text. Thus, we think that through our analytic method, we were able to validly categorise and assign quality ranks to the wiki contributions. Our main finding, that stands in line with previously reported findings on quantitative data (Heimbuch & Bodemer, 2018), is that a generally useful type of additional guidance for technology-enhanced learning may not be enough. Depending on the task, it can be necessary to consider adequate individual personality variables. In this case the Need for Cognitive Closure was valid research subject, since we were explicitly interested in the processing of ambiguous information and if people produce better learning and knowledge construction outcomes when guided either implicitly or explicitly. In future analyses, it might be interesting to consider the constellations of both categorisation systems, as outlined in Table 3, and the type of provided guidance in order to take a closer look at possible correlations between these two variables. In addition, the test persons in this study had the opportunity to leave comments about their contributions to the discussion and choice of topic. For future analyses, it may be interesting to determine how the selection of topics was justified and whether there was a correlation between the justification and the contribution to the discussion.

## References

- Bodemer, D., & Buder, J. (2006). Supporting collaborative learning with augmented group awareness tools. In R. Sun & N. Miyake (Eds.), *Proceedings of the Twenty-Eighth Annual Conference of the Cognitive Science Society* (pp. 77–82). Mahwah, NJ: Lawrence Erlbaum.
- Cole, M. (2009). Using Wiki technology to support student engagement: Lessons from the trenches. *Computers & Education*, 52(1), 141–146.
- Cress, U., & Kimmerle, J. (2007). A theoretical framework of collaborative knowledge building with wikis: A systemic and cognitive perspective. In *Proceedings of the 8th International Conference on Computer Supported Collaborative Learning* (pp. 156–164). New Brunswick, NJ: International Society of the Learning Sciences.
- Cress, U., & Kimmerle, J. (2008). A systemic and cognitive view on collaborative knowledge building with wikis. *International Journal of Computer-Supported Collaborative Learning*, 3(2), 105.
- Dijksterhuis, A., van Knippenberg, A., Kruglanski, A. W., & Schaper, C. (1996). Motivated Social Cognition: Need for Closure Effects on Memory and Judgment. *Journal of Experimental Social Psychology*, 32(3), 254–270.
- Dillenbourg, P. (2002). Over-scripting CSCL: The risks of blending collaborative learning with instructional design. *Three Worlds of CSCL. Can We Support CSCL?*, 61–91.
- Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis: a regression-based approach* (Second edition). New York: Guilford Press.

- Hayes, A. F., Montoya, A. K., & Rockwood, N. J. (2017). The analysis of mechanisms and their contingencies: PROCESS versus structural equation modeling. *Australasian Marketing Journal (AMJ)*, 25(1), 76–81.
- Heimbuch, S., & Bodemer, D. (2017). Controversy awareness on evidence-led discussions as guidance for students in wiki-based learning. *The Internet and Higher Education*, 33(2017), 1–14.
- Heimbuch, S., & Bodemer, D. (2018). Interaction of guidance types and the Need for Cognitive Closure in wiki-based learning. *PeerJ*, 6, e5541.
- Johnson, D. W., Johnson, R. T., & Smith, K. A. (2000). Constructive controversy: The educative power of intellectual conflict. *Change: The Magazine of Higher Learning*, 32(1), 28–37.
- Johnson, D. W., Johnson, R. T., & Tjosvold, D. (2000). Constructive controversy: The value of intellectual opposition. In M. Deutsch, P. T. Coleman, & E. C. Marcus (Eds.), *The handbook of conflict resolution: Theory and practice* (pp. 65–85). San Francisco, CA: Jossey-Bass.
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist*, 41(2), 75–86.
- Kump, B., Moskaliuk, J., Dennerlein, S., & Ley, T. (2013). Tracing knowledge co-evolution in a realistic course setting: A wiki-based field experiment. *Computers & Education*, 69, 60–70.
- Laugwitz, B., Held, T., & Schrepp, M. (2008). Construction and Evaluation of a User Experience Questionnaire. In A. Holzinger (Ed.), *HCI and Usability for Education and Work* (pp. 63–76). Springer Berlin Heidelberg.
- Majchrzak, A., Wagner, C., & Yates, D. (2006). Corporate wiki users: results of a survey. In *Proceedings of the 2006 international symposium on Wikis - WikiSym '06* (p. 99). Odense, Denmark: ACM Press.
- Mayring, P. (2014). *Qualitative content analysis: theoretical foundation, basic procedures and software solution*. Klagenfurt.
- Mayring, P. (2015). Qualitative Content Analysis: Theoretical Background and Procedures. In A. Bikner-Ahsbals, C. Knipping, & N. Presmeg (Eds.), *Approaches to Qualitative Research in Mathematics Education: Examples of Methodology and Methods* (pp. 365–380). Dordrecht: Springer Netherlands.
- McLoughlin, C. (2002). Computer supported teamwork: An integrative approach to evaluating cooperative learning in an online environment. *Australasian Journal of Educational Technology*, 18(2).
- Miller, M., & Hadwin, A. (2015). Scripting and awareness tools for regulating collaborative learning: Changing the landscape of support in CSCL. *Computers in Human Behavior*.
- Murray, M. H. (2003). Managing teamwork online. In *A forum for engaging, designing, assessing, and collaborating in online learning and teaching* (pp. 75–82). Brisbane, Australia: Queensland University of Technology.
- Papadopoulos, P. M., Demetriadis, S. N., & Weinberger, A. (2013). ‘Make it explicit!’: Improving collaboration through increase of script coercion. *Journal of Computer Assisted Learning*, 29(4), 383–398.
- Piaget, J. (1977). *The development of thought: equilibration of cognitive structures*. Viking Press.
- Ramsenthaler, C. (2013). Was ist „Qualitative Inhaltsanalyse?“. In M. Schnell, C. Schulz, H. Kolbe, & C. Dunger (Eds.), *Der Patient am Lebensende* (pp. 23–42). Wiesbaden: Springer Fachmedien Wiesbaden.
- Schlichter, J., Koch, M., & Bürger, M. (1998). Workspace awareness for distributed teams. In W. Conen & G. Neumann (Eds.), *Coordination Technology for Collaborative Applications* (Vol. 1364, pp. 199–218). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Schlink, S., & Walther, E. (2007). Kurz und gut: Eine deutsche Kurzsкала zur Erfassung des Bedürfnisses nach kognitiver Geschlossenheit. *Zeitschrift Für Sozialpsychologie*, 38(3), 153–161.
- Schumpe, B. M., Brizi, A., Giacomantonio, M., Panno, A., Kopetz, C., Kosta, M., & Mannetti, L. (2017). Need for Cognitive Closure decreases risk taking and motivates discounting of delayed rewards. *Personality and Individual Differences*, 107, 66–71. <https://doi.org/10.1016/j.paid.2016.11.039>
- Vertzberger, Y. Y. (1990). The world in their minds: Information processing, cognition, and perception in foreign policy decisionmaking.
- Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. *Journal of Personality and Social Psychology*, 67(6), 1049–1062. <https://doi.org/10.1037/0022-3514.67.6.1049>
- Webster, D. M., & Kruglanski, A. W. (1997). Cognitive and Social Consequences of the Need for Cognitive Closure. *European Review of Social Psychology*, 8(1), 133–173.
- Wichmann, A., & Rummel, N. (2013). Improving revision in wiki-based writing: Coordination pays off. *Computers & Education*, 62, 262–270.

## Acknowledgments

We would like to thank Manon-Lea Laudien, B.Sc. for preparing the data in the course of her Bachelors’ thesis.



# Expansively Framing Social Annotations for Generative Collaborative Learning in Online Courses

Christopher D. Andrews, Grant T. Chartrand, and Daniel T. Hickey  
andrewch@indiana.edu, gchartra@indiana.edu, dthickey@indiana.edu  
Indiana University

**Abstract:** This study extends the use of *expansive framing*, a discursive pedagogical practice for supporting generative (i.e., transferable) collaborative learning, to a fully online undergraduate Educational Psychology course. This study examined how students expansively framed their engagement in a social annotation activity across the semester. Quantitative analysis confirmed the extent to which interactions in the annotation activity were expansively framed and found a significant correlation between expansive framing and open-ended exam performance. Qualitative analyses confirmed that expansively framed interactions made numerous connections between disciplinary course knowledge and nascent disciplinary teaching practices. More generally, the study showed that expansive framing can be easily and successfully used to support generative collaborative learning in online courses.

## Introduction

Expansive framing (Engle, 2006; Engle, Lam, Meyer, & Nix, 2012) is a recent framework for explaining and encouraging generative learning—learning that is more likely to transfer to subsequent contexts. The research on transfer is plentiful and has traditionally focused on learners’ abilities to create abstract mental representations of the content to be transferred (Day & Goldstone, 2012). Engle (2006) argued that, from a situative perspective, “transfer is more likely to occur to the extent that learning and transfer contexts have been framed to create what is called *intercontextuality* between them” (p. 456).

Intercontextuality is created when learners make enough connections back and forth between a learning context and a transfer or imagined future context (Engle et al., 2012). This intercontextuality indicates that the social context—the who, when, where, how, and why—expands to encompass both the learning context and the transfer context. According to Engle (2006), “When this [intercontextuality] occurs . . . the content established during learning is considered relevant to the transfer context” (p. 456). If learners perceive the learning context and the transfer context to be relevant to each other, it is more likely that transfer will occur. Engle suggested that this intercontextuality can be created through what she referred to as *expansive framing*.

Expansive framing uses prompts or cues (usually from the instructor) that position learners as authors of their own ideas and encourage learners to frame their immediate learning context across time (relevance of prior learning to current learning context and relevance of current learning to potential future contexts), places, topics, and participants. These “meta-communicative signals” (Engle, 2006, p. 456) create intercontextuality by helping learners see prior learning as relevant to immediate learning contexts and the potential for current learning to be useful in other contexts and situations. Expansive framing pushes learners to (a) be publicly recognized as authors of the connections they are making to other learning contexts, (b) be held accountable for their contributions, and (c) feel encouraged to adapt and generate new connections. Expansive framing contrasts with *bounded* framing which (a) focuses on the contexts defined and/or provided by the instructor, (b) discourages learners from making connections or considering other contexts of use, and (c) expects learners to explain the ideas of a text or the teacher (Engle et al., 2012).

This study aims to extend the work of expansive framing in three innovative ways. First, studies of expansive framing have primarily focused on the instructors’ use of expansive framing in classrooms such as an elementary classroom (Engle, 2006), tutoring sessions (Engle et al., 2011), a secondary science classroom (Engle et al., 2012), and a hybrid second language undergraduate course (Mendelson, 2010). More recent research has acknowledged the role of the learner and examined whether learners’ perceptions and beliefs aligned with expansive framing (Lam, Mendelson, Meyer, & Goldwasser, 2014); however, few studies examined whether students took up expansive framing in their own discourse within a course (e.g., Fasso & Knight, 2015). While expansive framing was originally conceived as a *pedagogical* tool used by teachers to encourage learning, we reconceptualized expansive framing as a *learning* tool used by students to promote their own generative learning. This view also prioritizes students’ agency in taking up expansive framing for learning. One aim of this study is to examine how and to what extent *students* took up expansive framing.

Second, we argue that expansive framing is relatively easy to implement in asynchronous online settings and would likely support generative learning in those settings. A search of the literature returned few



instances of expansive framing in fully online settings. Fasso and Knight (2015) incorporated expansive framing into a study of adult learning in a hybrid a/synchronous professional development course for teachers from various contexts (e.g., P-12 education, vocational training, and higher education). Their study found that “expansive framing of the discussions led to broadened perspectives, enhanced collaboration and sharing, and evidence of the transfer of ideas across contexts that served to enrich the ideas of others” (p. 279). Hickey and Rehak (2013) and Jaber, Dini, Hammer, and Danahy (2018) successfully implemented elements of expansive framing in asynchronous online settings by encouraging learners to connect disciplinary knowledge to their relevant disciplinary practices; however, their studies were based off Engle’s prior work of productive disciplinary engagement (Engle & Conant, 2002). Therefore, a second aim of this study is to examine expansive framing in a completely asynchronous online setting.

Finally, this study extends expansive framing by combining it with the use of a social annotation tool. Annotation tools are well known in the CSCL community for supporting anchored discussions, or discussions anchored to an artifact such as a course reading (van der Pol, Admiraal, & Simons, 2006). Novak, Razzouk and Johnson’s (2012) literature review found mixed results from using social annotation tools for learning. Most of those studies, however, focused on attitudes rather than learning outcomes. Social annotation has been shown to foster elaborated discussions (Eryilmaz, van der Pol, Ryan, Clark, & Mary, 2013; Gao, 2013). Social annotation aligns with expansive framing because it allows learners to layer their own expansively framed contexts directly on top of content discussions and readings. The third aim of this study examines whether social annotation tools can support the use of expansive framing.

## Research questions and methods

This study asked (1) to what extent were students’ interactions expansively framed, (2) how was expansive framing related to individual learning outcomes, and (3) what was it about expansively framed interactions that appeared generative.

## Context and participants

This study took place in an online undergraduate Educational Psychology course for pre-service teachers which took place in the Spring 2018 semester at a large Midwestern university. The first author was the lead instructor for this course. There were 17 student participants from four different majors (visual arts education, n=10; world language education, n=4; physical education, n=2; dietetics, n=1). During the 16-week course, students annotated 23 course readings using the online social annotation tool *Hypothesis* (<https://web.hypothes.is/>). Hypothesis allows students to select portions of text and directly annotate it by adding their own comments. Hypothesis allows users to engage in further conversation via threaded comments on the direct annotations. The course readings consisted of academic articles that discussed key theoretical principles from several major learning theories. The students were asked to engage in the annotation activity but were not graded on their participation in the annotations. Students were, however, asked to reflect on peers’ and their own annotations on a weekly reflection which was graded. The prompt for engaging in the annotation activity was deliberately worded to encourage students to expansively frame their engagement and their annotations (see Table 1).

Table 1: Annotation activity prompt

---

Before you start reading, you should think about a specific context for your engagement in the reading and annotations. To do this, you should:

1. Pick a developmental level (e.g., Elementary, Middle, Secondary, etc.) you are interested in teaching.
2. Pick a content standard(s) from the State Standards that you are interested in focusing on.

As you read and make web annotations, consider the relevance of the reading and annotations as applied to your context. At the end of most units, you will be designing a lesson plan that you could actually use in a current or future classroom as part of our Theory/Practice Workshops. Picking a developmental level and content-area standard(s) should help you think about how what you are learning applies to the Theory/Practice Workshops as well as other potential future contexts (e.g., classroom, student teaching, other classes, etc.).

In this class we will be using web annotations using a web tool called Hypothes.is to participate in a larger discussion of articles that we will read as well as facilitate making connections to other content such as other articles we are reading in the class or content you may read or interact with outside of class. The web annotations can also help us make connections to prior experiences you may have had, current experiences you are having, or potential future experiences you may have (e.g., your field experience, future classroom, or work-related experiences).

---

At the end of the course, students took a written exam that included two different prompts relating to learning theory. One prompt provided students with an op-ed article written in the New York Times arguing for the pedagogical practice of lecturing. Students were asked to analyze that article from two different theoretical perspectives of their choice. For each theory, students needed to (1) describe what learning looks like in that theory, (2) describe a specific concept or aspect of that theory, (3) explain how that theory and concept would disagree with the article, and (4) provide an appropriate link to the course readings to back up their argument. The second prompt asked students to (1) design an assessment and two activities for a specific academic standard in their content area provided for them by the instructor, (2) describe how a particular learning theory informed their design, and (3) provide an appropriate link to the course readings. The exam was designed to assess students' understanding of the major learning theories discussed in the course and their ability to apply it to classroom design. Within the constraints of a typical teacher education course, this exam was a reasonable proxy for a transfer measure that was indicative of generative learning and was likely to transfer to subsequent educational and professional contexts.

Table 2: Coding scheme for aspects of expansive framing

| Code                  | Description   | Example   |
|-----------------------|---|---|
| <b>Time Past</b>      | Refers to a past time that what was learned then can, should be, or has been used.  | “I had an English teacher senior year...”<br>“This reminded me of last semester...”                     |
| <b>Time Future</b>    | Refers to a future time in which what is learned can, should be, or would be used.  | “In my future classroom I’d like to...”<br>“Teachers need to...”  |
| <b>Place</b>          | Refers to another place outside of the course in which what is learned can, should be, or has been used.  | “In my field experience...”<br>“At the high school...”<br>“At home...”                                  |
| <b>Topic</b>          | Refers to a non-course topic in which what is learned can, should be, or has been used.   | “In my content area of visual arts...”<br>“My Pluralism in Education course talked about this...”       |
| <b>Participants</b>   | Refers to person(s) other than the teacher that one can, should be, or has communicated what one has learned.   | “Pre-service teachers could use this to...”<br>“This really applies to students...”                     |
| <b>Accountability</b> | Holds others accountable for sharing knowledge by directly engaging others with questions to encourage threaded discussion and/or responding directly to a peer's statement by answering and/or referring to a specific question/comment. | “How would you...?”<br>“Leslie’s comment helped me think about...”<br>“I agree with your claim that...” |
| <b>Authorship</b>     | Presents themselves as authoring knowledge.   | “I think this is important...”<br>“We also should consider...”  |

## Data collection and analysis

All of the annotations during the course of the semester were exported from Hypothesis. This data included the text of students' annotations, the annotated text from the article (if applicable), the level of the annotation (for threaded comments), and other metadata.

To address the first research question, these annotations were first coded for enlistment of expansive framing using the coding scheme in Table 2 following from Engle et al. (2012). This coding revealed the proportion of annotations that were expansively framed with each aspect. To determine how expansive those annotations were, we developed a second coding scheme (Table 3). For this degree of expansiveness coding scheme, the first two authors coded 20% of the annotations separately and achieved an inter-rater reliability of 0.70 using Cohen's Kappa. The first author then coded the remaining annotations alone.

To address the second research question and determine whether students' engagement in expansive framing appeared to be generative to a transfer task, we ran a Pearson correlation between students' mean expansive framing score (i.e., degree of expansiveness) and student scores on the written final exam. Two students dropped the course before the final exam and thus were not included in this analysis. To address the third research question, we drew upon discourse analysis (Wooffitt, 2005) to identify what students were actually doing in their annotations that might lead to generative learning. Looking across the dataset, we were

interested in patterns that emerged around how students used framing in their annotations and how that framing shaped the subsequent discourse in the threaded annotations.

Table 3: Coding scheme for degrees of expansiveness

| Code                                    | Description   | Example   |
|---|---|---|
| <b>U</b><br><b>Unframed</b>             | The annotation makes no reference to any aspect of expansive framing.   | “I’m not sure I fully understand the difference between the two.”   |
| <b>0</b><br><b>Bounded</b>              | The annotation references at least one aspect of expansive framing, but the annotation is confined to the course and content from the current reading.  | “Further analysis and application of what is learned can occur when knowledge of subject matter exists. How the subject matter is organized and connects to itself matters.”  |
| <b>1</b><br><b>Slightly Expansive</b>   | Uses <i>at least 1 aspect</i> of expansive framing; Uses <i>vague descriptions or connections</i> to aspects of expansive framing; AND/OR <i>Does not go beyond</i> the framing established by the original document or annotation it responded to. | “I think the things that they learn in school will have an effect on their lives outside of school, it just depends on the student and how much of the information they retain. I do agree though that the teachers should try to have an understanding of the students' lives outside of school.”  |
| <b>2</b><br><b>Moderately Expansive</b> | Uses <i>more than 1 aspect</i> of expansive framing; Uses <i>specific examples</i> when connecting to aspects of expansive framing; OR <i>Explicitly orients towards others</i> in a future setting.  | “In high school, I had a Spanish teacher that taught English in Spain for about 7 years. She was fluent in Spanish so I thought I would learn a lot during my time in her class. It was the complete opposite, which is interesting to me because of this highlighted portion. Just because she was an expert on the topic, she could not teach it well.”   |
| <b>3</b><br><b>Very Expansive</b>       | Uses <i>more than 2 aspects</i> of expansive framing; Uses <i>specific examples</i> when connecting to aspects of expansive framing; AND <i>Explicitly orients towards others</i> in a future setting.  | “This is a great way to engage students in their learning and understanding of art history. When analyzing a painting, students must first understand how to use contextual clues to describe and explain what the painter might have been communicating through his/her work. Taking a step further, to help students improve their flexibility in transferring knowledge, a teacher might ask, ‘What if the figure on the right was looking toward the viewer, rather than away. What might we say about the painting then?’” |

## Results

The first major finding was that students did indeed reference aspects of expansive framing in their discourse (see Table 4). Across the course, students generated a total of 459 annotations. Students referenced Participants most often (76% of student annotations), followed by Authorship (71%) and Topics (61%). The types of Participants referenced most often were teachers and students. The types of Topics referenced varied but were generally related to three different categories: their own content area, other courses they were enrolled in, or general education topics and activities. Students did not reference all aspects of expansive framing equally, however. Students did not reference Places (28%) or Accountability (39%) as often. This shows that students have expansively framed their discourse, favoring some aspects over others.

The coding results of degree of expansiveness provides further evidence that students used expansive framing in their discourse. Only nine (2%) of the 459 annotations were unframed and 12 (3%) were bounded to the course and course reading. This means that more than 95% of the annotations were at least slightly expansive and 62% of the annotations were coded as moderately or very expansive.

The Pearson correlation between students average score on expansive framing (degrees of expansiveness) and their written final exam score was 0.56 ( $p < .05$ ). In other words, students who performed better on the final written exam also, on average, were more expansive in their annotations. The final exam assessed students’ ability to analyze educational practices from a theoretical perspective as well as design

towards educational practices grounded in theory. These results were promising but warranted a qualitative exploration to examine how students used expansive framing for generative learning.

Table 4: Coding results of students' use of expansive framing in their annotations

| Code                         | #   | %    |
|------------------------------|-----|------|
| Aspects of expansive framing |     |      |
| Time Past                    | 179 | 39%  |
| Time Future                  | 227 | 50%  |
| Place                        | 129 | 28%  |
| Topic                        | 279 | 61%  |
| Participants                 | 349 | 76%  |
| Accountability               | 179 | 39%  |
| Authorship                   | 324 | 71%  |
| Degree of expansiveness      |     |      |
| U (Unframed)                 | 9   | 2%   |
| 0 (Bounded)                  | 12  | 3%   |
| 1 (Slightly expansive)       | 152 | 33%  |
| 2 (Moderately expansive)     | 203 | 44%  |
| 3 (Very expansive)           | 83  | 18%  |
| Total                        | 459 | 100% |

Discourse analysis of the annotations revealed several patterns that appeared both productive (Engle & Conant, 2002) and generative. One such pattern was that students were connecting disciplinary practices in their own content areas to the disciplinary knowledge in the course. Students made connections to their own content areas in all but four of the course readings (19 of 23) and in 92 annotations (20% of all annotations). The example below (see Table 5) shows a threaded conversation between students on an article about assessment.

Table 5: Threaded annotations connecting content-area disciplinary practices to course content (Excerpt 1)

---

*Anchor text from Shepard (2000):*

Students also reported that they had to be more honest about their own work as well as being fair with other students, and they had to be prepared to defend their opinions in terms of the evidence. (p. 12)

---

1. *Leslie, direct annotation (coded as moderately expansive)*

This is directly applicable to art critiques where students must analyze and evaluate the work of their classmates, and give them feedback. They must also receive feedback on their own work and be prepared to respond to this feedback (which requires self-assessment).

2. *Bethany (coded as moderately expansive)*

I feel like this can also apply to students studying a foreign language, if they are speaking to another student that is more skilled at the language than they are, then they can learn more and get assistance if they are struggling on vocabulary.

---

In this example, the original article was discussing implications of self-assessment practices and Leslie, a visual arts education student, points out a “direct” connection to her own content area and provides an example of a disciplinary practice in her content area. Likewise, Bethany, in her reply to Leslie, sees an application to students in her content area of foreign language learning. Both students expansively framed their learning by recognizing the relevance of prior knowledge from their own content area and made productive generalizations from the content to disciplinary practices in their own content area. Additionally, Bethany’s response suggests she read Leslie’s connection to art (“this can *also* apply”, emphasis added). Connecting to one’s own relevant contexts and being exposed to other relevant contexts likely promotes productive and generative collaborative learning.

Another pattern was students making connections to general disciplinary contexts and practices of teachers. This pattern was found in all but one of the course readings (22 of 23) and in 156 annotations (34% of all annotations). Table 6 shows a threaded conversation about deficit approaches to education.

Table 6: Threaded annotations connecting general disciplinary practices to course content (Excerpt 2)

---

*Anchor text from Paris (2012):*

Simply put, the goal of deficit approaches was to eradicate the linguistic, literate, and cultural practices many students of color brought from their homes and communities and to replace them with what were viewed as superior practices. (p. 93)

---

1. *Rose, direct annotation (coded as moderately expansive)*

In one of my other classes, we are learning about Native American education. One thing that I have learned is that by eradicating student's culture, can make them feel unimportant and unvalued. I think this is a really important idea to keep in mind in a future classroom because students will not feel motivated or engage if they feel like their voice is not important.

2. *Jamie (coded as very expansive)*

I agree with you and in my own experiences, I have seen school trying to do special holidays related to China such as the (Chinese New Year), but after talking about it with my friends from China, they all say that the school got some of the major information incorrectly. For example, there is no such thing as fortune cookie and lion dance in mainland China on Chinese New Year. Although it is great to see that American schools are embracing different cultures from other countries, it is also important to keep in mind as future teachers that we get the facts correct before presenting to the students to avoid conflicts between teacher and student.

3. *Instructor (not coded)*

As I mentioned in the mini-lecture, this is very similar to the difference approaches to education where we try to incorporate other cultures but in a way that "fits into" the dominant culture (which is why there are so many inaccuracies as you mentioned). So how do we value our students ways of doing and knowing and being in our classrooms in more culturally sustaining ways?

4. *Brittney (coded as very expansive)*

In one of my other courses (pluralism in education), we discussed this and similar questions. One of the best solutions we thought of was to bring in people, that held a certain view point or background, to talk to students. Bringing in people who have first hand experience, in a culture for example, allows the students to feel valued if they are from that culture and the others to learn about them in an authentic way.

5. *Marrin (coded as moderately expansive)*

I am in class with Brittney and took a lot away from this topic in class. We also talked about how bringing in an outside resource can help bring a different learning experience than a daily teacher can lead. By simply having a new face leading sometimes allows students to be more engaged but also different methods of teaching is important in some instances. Another aspect we talked about is that a specific person teaching a subject can go a lot further than what the teacher may be. For example, if a white male were to be teaching a lesson about american indians students may not be as interested because he obviously would not be a direct source of their culture, though if an american indian were to come into class and teach a lesson students would value the information he shared much more.

---

In this annotation thread, Rose connects the content of the reading to another course she is taking and views her learning as relevant to the disciplinary practices of "a future classroom." Jamie responds by sharing an experience that validates Rose's prior learning. Jamie concludes her response by explicitly orienting towards herself and others in the class ("as future teachers . . . we") in sharing an example of how this applies to general disciplinary practices ("get the facts correct" about cultural activities). This type of comment was not uncommon throughout the annotations as students often used the phrase, "we need to" or "as future teachers we should" when writing about future applications of the content for students and teachers. The instructor of the course (the first author) responds to Jamie and connects back to the course content, followed by a question that asks students to imagine how they might apply the content of the article to their own future students. Brittney replies to the instructor's question by providing an example from another course she is enrolled in. Marrin states she is "in class with Brittney" suggesting that her response is not oriented toward Brittney, but rather towards others in the class. Marrin then reiterates Brittney's suggestion, validating her contribution, and connects to general disciplinary practices by briefly suggesting that "different methods of teaching" may be helpful. Finally,

Marrin shares an additional example of a disciplinary practice that connects to the previous responses. This set of threaded annotations showed students sharing examples of prior learning that was relevant to the content they were engaging in and generalizing the content to their future disciplinary practices.

A final example (see Table 7) demonstrates a pattern that was less common—found in 18 of 23 course readings and 46 annotations (10%)—but was productive for student’s engagement and likely to be generative. Most of the annotations that were coded for accountability were simply students responding to a peer’s annotation. However, some of the most productive responses occurred when students requested ideas from their peers. These questions often led to students sharing potential disciplinary practices from their own content area or for general education practices.

Table 7: Threaded annotations stemming from a student question (Excerpt 3)

---

*Anchor text from Ito et al. (2013):*

Further, when individual competence is assessed based on grades, test scores, and other standardized and summative metrics, one student’s success highlights another student’s failure. Environments like the HPA, Quest to Learn, or Clarissa’s online writing group have a different dynamic (p. 48)

---

1. *Anna, direct annotation (coded as very expansive)*

How can we create this in a classroom? What can we as teachers do to create an environment where one student's success does not highlight another's failure, because standardized tests and evaluative exams will be inevitable. How can we create a group of students that genuinely enjoys collaborating and wants to do well for the good of the class, and not themselves, when we will undeniably have tension between individual students because of personality difference and difference in interest? **Please respond w ideas: I'm genuinely interested in how we can do this!** (emphasis in original)

2. *Abby (coded as very expansive)*

These are great questions that I'm wondering about myself also! For your first Q "What can we as teachers do to create an environment where one student's success does not highlight another's failure?" One example that popped into my head is how during many classes I took during high school (and even in college) the teacher/prof would state "the highest grade was a 98 and the lowest was a 32" or something like that. I feel like this isn't appropriate to do, especially in a high school classroom where it is VERY easy to find out which student it was that the highest or lowest score belonged to. I feel like announcing the scores out loud kind of pits the students against one another, and highlights failures. Instead of the teacher announcing the scores, they should talk to the struggling students privately and see what needs to be done to bring their grades up. There's no point in discouraging the students by stating how much better another student did compared to them. That's just one example for the first question, but I'd also like to hear more from others! (that goes for the rest of the questions, too)

3. *Brittney (coded as moderately expansive)*

Personally I have found goal setting to be a great way to motivate students without creating a competitive environment. I'm currently in a 1 credit swimming class, and at the beginning of the semester the instructors asked us to write down several goals we wanted to accomplish. I thought this activity helped make the class more personal to all of us and helped us appreciate when we or others accomplished their goals. This task would obviously look very different in a school classroom, but I think there are always ways to apply goal setting to our individual content areas.

---

In this example, Anna asks a series of questions on how to implement the ideas from the original article to their own potential future educational practices. At the end of her questions Anna emphasizes her request by using bold font to catch the attention of her peers and elicit their responses. Anna is promoting the authorship of her peers and holding them accountable for their ideas by requesting them to share. In Abby’s response, she recalls a bad example from a prior situation (highlighting students’ failures) and suggests a general alternative practice (talking with struggling students privately). Abby ends her annotation with a request for more responses (“I’d also like to hear more from others!”). Brittney’s response shares a prior experience that is relevant (setting goals) to the questions Anna asked and suggests that this practice could be applicable in anyone’s content area. Holding students accountable for their own ideas by asking questions, although not used as often as other framing aspects, encouraged generative learning.

## Discussion and implications

This study provided initial evidence that online course readings can be expansively framed, that students can collaboratively take up expansive framing in online annotations, and that expansive framing appears to result in more generative learning. While acknowledging that a course with pre-service teachers may have been particularly conducive, expansive framing via Hypothesis enabled these students to collaborate in productive threaded conversations by connecting content-specific and more general disciplinary practices to the disciplinary knowledge presented in course readings. As Engle and colleagues (2006; 2012) have argued, these connections encourage *generative* learning which will likely transfer to future educational, personal, and professional contexts. This study showed that expansive framing is possible in online settings. Additionally, the prompt (see Table 1) could easily be modified to function in other similar online courses, especially pre-professional courses where students are working towards specific future contexts.

## References

- Day, S. B., & Goldstone, R. L. (2012). The import of knowledge export: Connecting findings and theories of transfer of learning. *Educational Psychologist, 47*(3), 153–176.
- Engle, R. A. (2006). Framing interactions to foster generative learning: A situative explanation of transfer in a community of learners classroom. *The Journal of the Learning Sciences, 15*(4), 451–498.
- Engle, R. A., & Conant, F. R. (2002). Guiding principles for fostering productive disciplinary engagement: Explaining an emergent argument in a Community of Learners classroom. *Cognition and Instruction, 20*(4), 399–483. [http://doi.org/10.1207/S1532690XCI2004\\_1](http://doi.org/10.1207/S1532690XCI2004_1)
- Engle, R. A., Lam, D. P., Meyer, X. S., & Nix, S. E. (2012). How does expansive framing promote transfer? Several proposed explanations and a research agenda for investigating them. *Educational Psychologist, 47*(3), 215–231. <http://doi.org/10.1080/00461520.2012.695678>
- Engle, R. A., Nguyen, P. D., & Mendelson, A. (2011). The influence of framing on transfer: Initial evidence from a tutoring experiment. *Instructional Science, 39*(5), 603–628.
- Eryilmaz, E., van der Pol, J., Ryan, T., Clark, P. M., & Mary, J. (2013). Enhancing student knowledge acquisition from online learning conversations. *International Journal of Computer-Supported Collaborative Learning, 8*(1), 113–144.
- Fasso, W., & Knight, B. A. (2015). Knowledge transfer in community-embedded learning: A case study. *Journal of Education Research, 9*(3), 267–281.
- Gao, F. (2013). A case study of using a social annotation tool to support collaboratively learning. *Internet and Higher Education, 17*(1), 76–83. <http://doi.org/10.1016/j.iheduc.2012.11.002>
- Hickey, D. T., & Rehak, A. M. (2013). Wikifolios and participatory assessment for engagement, understanding, and achievement in online courses. *Journal of Educational Multimedia and Hypermedia, 22*, 407–441.
- Ito, M., Gutiérrez, K., Livingstone, S., Penuel, B., Rhodes, J., Salor, K., ... Watkins, S. C. (2013). *Connected learning: An agenda for research and design*. Irvine, CA: Digital Media and Learning Research Hub.
- Jaber, L. Z., Dini, V., Hammer, D., & Danahy, E. (2018). Targeting disciplinary practices in an online learning environment. *Science Education, 102*(4), 668–692. <http://doi.org/10.1002/sc.21340>
- Lam, D. P., Mendelson, A., Meyer, X. S., & Goldwasser, L. (2014). Learner alignment with expansive framing as a driver of transfer. In J. L. Polman, E. A. Kyza, D. K. O'Neill, I. Tabak, W. R. Penuel, A. S. Jurov, ... L. D'Amico (Eds.), *The International Conference of the Learning Sciences* (pp. 689–696). Boulder, CO: International Society of the Learning Sciences.
- Mendelson, A. (2010). Using online forums to scaffold oral participation in foreign language instruction. *L2 Journal, 2*(1), 23–44. Retrieved from <https://escholarship.org/uc/item/8xs1r2tq>
- Novak, E., Razzouk, R., & Johnson, T. E. (2012). The educational use of social annotation tools in higher education: A literature review. *Internet and Higher Education, 15*(1), 39–49.
- Paris, D. (2012). Culturally sustaining pedagogy: A needed change in stance, terminology, and practice. *Educational Researcher, 41*(3), 93–97. <http://doi.org/10.3102/0013189X12441244>
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher, 29*(7), 4–14. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1002/9780470690048.ch10/summary>
- Wooffitt, R. (2005). *Conversation analysis and discourse analysis: A comparative and critical introduction*. London, UK: SAGE.
- van der Pol, J., Admiraal, W., & Simons, P. R. J. (2006). The affordance of anchored discussion for the collaborative processing of academic texts. *International Journal of Computer-Supported Collaborative Learning, 1*(3), 339–357. <http://doi.org/10.1007/s11412-006-9657-6>

# Unpacking Collaborative Learning Processes During Hands-on Activities Using Mobile Eye-Trackers

Bertrand Schneider, Harvard University, [bertrand\\_schneider@gse.harvard.edu](mailto:bertrand_schneider@gse.harvard.edu)

**Abstract:** This paper describes a multimodal dataset captured during a collaborative learning activity typical of makerspaces. Participants were introduced to computational thinking concepts using a block-based environment: they had to program a robot to solve a variety of mazes. Mobile eye-trackers, physiological wristbands and motion sensors captured their behavior and social interactions. In this paper, I leverage prior work on joint visual attention (Tomasello, 1995) and analyze the eye-tracking data collected during the study. This paper provides three contributions: 1) I use an emerging methodology to capture joint visual attention in a co-located setting using mobile eye-trackers (Schneider & al., 2018); 2) I then replicate findings showing that levels of joint visual attention are positively correlated with collaboration quality; 3) finally, I present a new measure that captures cycles of collaborative / individual work, which is positively associated with learning gains (but not with collaboration quality). I discuss these results and conclude with implications for capturing students' interactions in co-located spaces using Multimodal Learning Analytics.

## Introduction

In the last decade there has been a growing interest in cultivating skills that are not traditionally taught in traditional school settings. Those skills are often referred to as “21st century skills” (Dede, 2010; e.g., Collaboration, Communication, Creativity, Critical Thinking) because they are deemed essential for jobs that do not yet exist. New learning environments, such as digital fabrication labs and makerspaces, are ideal spaces for their development. They allow students to learn complex concepts in STEM (Science, Technology, Engineering, Mathematics) through hands-on learning and applied projects. Measuring the development of those skills and providing formative assessment, however, remains a challenge (Berland, Baker, Blikstein, 2014), because each student is unique, and the development of those 21st century skills takes different forms depending on interacting factors (e.g. learners' personalities, prior knowledge, SES background). The CSCL community has long been studying those skills before they gained a renewed attention from researchers and the general public.

For the scope of this paper, I focus on students' collaboration and communication by leveraging a new field of research called Multimodal Learning Analytics (Blikstein & Worsley, 2016; MMLA) to capture the quality of learners' interactions. MMLA uses multiple high-frequency sensors to capture users' behavior and applies data mining techniques to find trends and predictors in large datasets. Joint visual attention has been extensively studied by social and developmental psychologists and has been shown to be critical to many social interactions. Based on prior literature (e.g., Richard & Dale, 2015; Schneider & Pea, 2013), the main hypothesis of this paper is that productive groups exhibit higher levels of joint attention compared to less productive groups. This construct was captured using multiple mobile eye-trackers in co-located spaces (Schneider & al., 2018). More specifically, I designed a hands-on task typical of makerspaces (i.e., learning to program a robot to solve a variety of mazes) and computed measures of joint visual attention. I correlated them with three outcomes measures: the quality of their collaboration (coded with a validated rating scheme in the learning sciences), their task performance (i.e., how successful they were) and their learning gains (computed from a pre and post-test). Finally, because collaboration can be a powerful way to support learning, I analyzed the eye-tracking data to find behaviors that were not just related to collaboration quality, but also learning outcomes.

This paper is structured as follows: the first part reviews the literature on dual eye-tracking and the various measures that researchers have developed over the years to capture joint visual attention. The second part describes the study, participants and data collection protocol. The third part discusses the steps to pre-process the data and compute metrics of joint visual attention to correlate them with outcomes of interests. Finally, I discuss our results and conclude with future steps for capturing students' 21st skills in makerspaces using MMLA.

## Literature review

This section provides a succinct review of foundational work in developmental and social psychology, as well as in Computer-Supported Collaborative Learning (CSCL) and Computer-Supported Collaborative Work (CSCW) where multiple eye-trackers are used to look at participants' visual alignment.

There are currently three strands of research studying collaborative processes through dual eye-tracking. The first strand uses remote eye-trackers to capture joint visual attention, where users are each looking at a different computer displays (for example through video conferencing). In an early study, Richardson & Dale (2005) explored the coupling between speakers' and listeners eye movements and its relationship with discourse



comprehension. They found a positive correlation between discourse comprehension and dynamic coupling between conversants' eye movement. In a subsequent study, they replicated those results for a live conversation (Richardson, Dale & Kirkham, 2007). In CSCL, researchers have used this methodology to study pair programming tasks (Jermann, Mullins, Nüssli & Dillenbourg, 2011) and found that collaboration quality was characterized by higher levels of "gaze cross-recurrence" (i.e., joint visual attention). A second strand of research has started to study more ecological settings using mobile eye-trackers. Yu and Smith (2013), for example, used mobile eye-trackers to explore infant cross-situational word learning through eye-hand coordination. In education, Schneider & al. (2018) studied apprentices in logistics interacting with a tangible user interface. They found that levels of joint visual attention (as captured by mobile eye-trackers) were correlated with their quality of collaboration. Additionally, they developed a methodology to capture leadership behaviors from dual eye-tracking data by identifying who initiated and who responded to an offer of joint visual attention. Imbalances of these behaviors were negatively correlated with learning gains. Finally, a last strand of research has started to explore the benefits of novel visualization techniques to improve learning in a collaborative setting, for example by displaying participants' gaze to each other. This intervention is sometimes called a "gaze awareness tool", "shared gaze visualization" or "Bidirectional Gaze" (for a review, see D'Angelo & Schneider, 2018). D'Angelo & Begel (2017) have enhanced remote pairs' speed and success in communicating when resolving a coding problem by using eye tracking devices to show each participant where their partner is looking on the screen. In education, Schneider & Pea (2013) found that making the gaze of each partner visible promoted interactions of higher quality and consequently increased students' learning gains.

In conclusion, there is ample work showing that joint visual attention is a central mechanism by which group members coordinate their actions and establish a common ground (Clark & Brennan, 1991). Furthermore, recent research has been leveraging new sensing technology to quantify joint visual attention in dyads of users (e.g., Jermann, Mullins, Nüssli & Dillenbourg, 2011). While most studies have looked at remote collaborations, there is some nascent work in co-located settings using mobile eye-trackers (e.g., Yu & Smith, 2013). Ultimately, however, the goal from a CSCL perspective is to understand how collaborative processes *contribute* to learning. Joint visual attention (JVA), for example, is a necessary but not sufficient condition for productive social interactions. This paper is about going beyond capturing JVA and finding more precise indicators of collaborative learning. This paper builds upon prior findings (e.g., Schneider & al., 2018), replicates results, and provides new contributions by isolating collaborative learning processes from the eye-tracking data.

## Methods

### Summary of the study

In this study, participants with no prior programming knowledge were given 30 minutes to program a robot to autonomously solve a series of increasingly complex mazes (see Fig. 1 for the setup of the experiment). Two different interventions were developed and used to support collaboration: a visualization of relative verbal contributions of the participants shown in real time and a brief informational explanation delivered verbally summarizing literature findings on the value of collaboration for learning. While dyads completed the activity, a variety of sensors described in 2.2 collected eye gaze, movement, verbal, and electrodermal activity data on participants. Dependent measures were an assessment of the quality of the collaboration, how well the participants coded the robot to perform the assigned task, and learning gains related to computational thinking. The study is described in more detail in Starr, Reilly & Schneider (2018).

### Participants

Participants were drawn from an existing study pool at a university in the northeastern United States. 42 pairs of participants (N=84) were used in the analysis. 62% of participants identified as students, with ages ranging from 19 to 51 years old (mean age = 26.7 years). 60% of participants identified as female. Participants were paid \$20 for the 90-minute session and did not know each other prior to the study.

### Experimental design

The study utilized a two-by-two between-subjects design where dyads were assigned to one of four conditions that would receive different interventions. 25% of dyads received neither intervention (Condition #1), 25% received solely the visualization intervention (#2), 25% received solely the informational intervention (#3) while the remaining quarter received both interventions (#4). The speech equity visualization utilized speech collected by the sensors in the experiment to display how much each participant spoke as a proportion of total talk during the activity. Dyads with this intervention saw a tablet display representing this data over the past 30 seconds by presenting colored rectangles that grew to take up more of the screen as relative contribution increased. The informational intervention involved a researcher reading a short passage that reminded dyads that they were

expected to collaborate and invited dyads to think about how they were collaborating during the activity. They were also told that research has found that equity of each partner's speech time is predictive of the quality of collaboration and learning gains. For an analysis of the differences between each experimental condition, please see Starr, Reilly & Schneider (2018).

## Procedure

After taking the pre-survey and calibrating all sensors, participants were shown a short tutorial video that introduced the basics of writing a program in Tinker, a block-based programming language designed for use with the microcontroller of the robot. Participants were then given five minutes to write code that would move the robot forward across a red line roughly two feet directly in front of it. The robot consisted of a microcontroller, two DC motors, and three proximity sensors. Following this tutorial activity, a second tutorial video was shown that highlighted more advanced features of Tinker such as using prewritten functions to turn the robot and using sensor values to trigger conditional statements. Dyads were also given a reference sheet summarizing the content covered in the tutorial video. Dyads then had 30 minutes to write code to navigate a robot through a series of mazes. Once the robot successfully completed a maze twice, a more challenging maze was provided. Dyads did not know the layout of the mazes ahead of time and were encouraged to write code that would allow the robot to solve any simple maze. During this portion of the study, the researcher provided standard hints at 5-minute intervals to all dyads regarding common pitfalls researchers identified in pilot testing of the activity.

## Independent, dependent measures and process data

The quality of the dyad's collaboration and task performance was assessed during the task by the researcher running the session. The quality of collaboration was measured by aggregating the nine scales adapted from Meier, Spada, & Rummel (2007): sustaining mutual understanding, dialogue management, information pooling, reaching consensus, task division, time management, technical coordination, reciprocal interaction, and individual task orientation (refer to Meier, Spada, & Rummel, 2007 for a definition of those terms). Researchers double-coded 20% of the sessions and had a Cronbach's alpha of .65 (75% agreement). Task behaviors evaluated included task performance (how many mazes were completed by the robot in 30 minutes), task understanding (how much major coding concepts were included in the design such as using sensors with appropriate thresholds in conditional statements), and improvement over time (how much a team's understanding of requisite technical skills and conceptual understanding of the task changed over time). The final written code of the dyad's was also assessed to determine theoretically how well it could have performed the assigned task barring technical issues. To assess learning of computational thinking principles, participants individually completed pre- and post-surveys consisting of four questions related to conditional statements, looping, and interpreting the output of given code (adapted from Brennan & Resnick, 2012; Weintrop & Wilenski, 2015). After the activity, participants also self-assessed their collaboration and wrote a brief reflection regarding how their thinking changed over the course of the activity. Table 1 presents a summary of the measures described in this section.

Table 1: Independent, process and dependent measures looked at in this paper (described above)

| Independent Measures (2x2)    | Process Measures                                       | Dependent Measures   |
|-------------------------------|--|--|
| Speech visualization (on/off) | Eye-tracking data:                                     | Collaboration quality (9 sub-scores; 1 overall score)          |
| Verbal intervention (yes/no)  | Individual gaze points on AOIs, Joint Visual Attention | Task performance: code quality, improvement, # of mazes solved |
|                               |  | Learning gains (computational thinking)                        |

## Dual eye-tracking measures and hypotheses

Prior work has explored multiple ways of capturing joint visual attention and collaborative processes from dual eye-tracking data. In this paper, I first follow a methodology described by Schneider & al. (2018) to compute joint visual attention from mobile eye-trackers and attempt to replicate previous results showing that JVA is associated with collaboration quality. Second, I was inspired by previous results showing that collaborative problem-solving is a cycle between moments of understanding and non-understanding (Miyake, 1986), and that ideal cycles of communication are related to group performance (Tschan, 2002). In this paper, I hypothesize that collaborative learning interactions are characterized by more frequent cycles of individual work and group interactions – which are captured from the eye-tracking dataset. In short, the hypotheses of this paper are as follows:

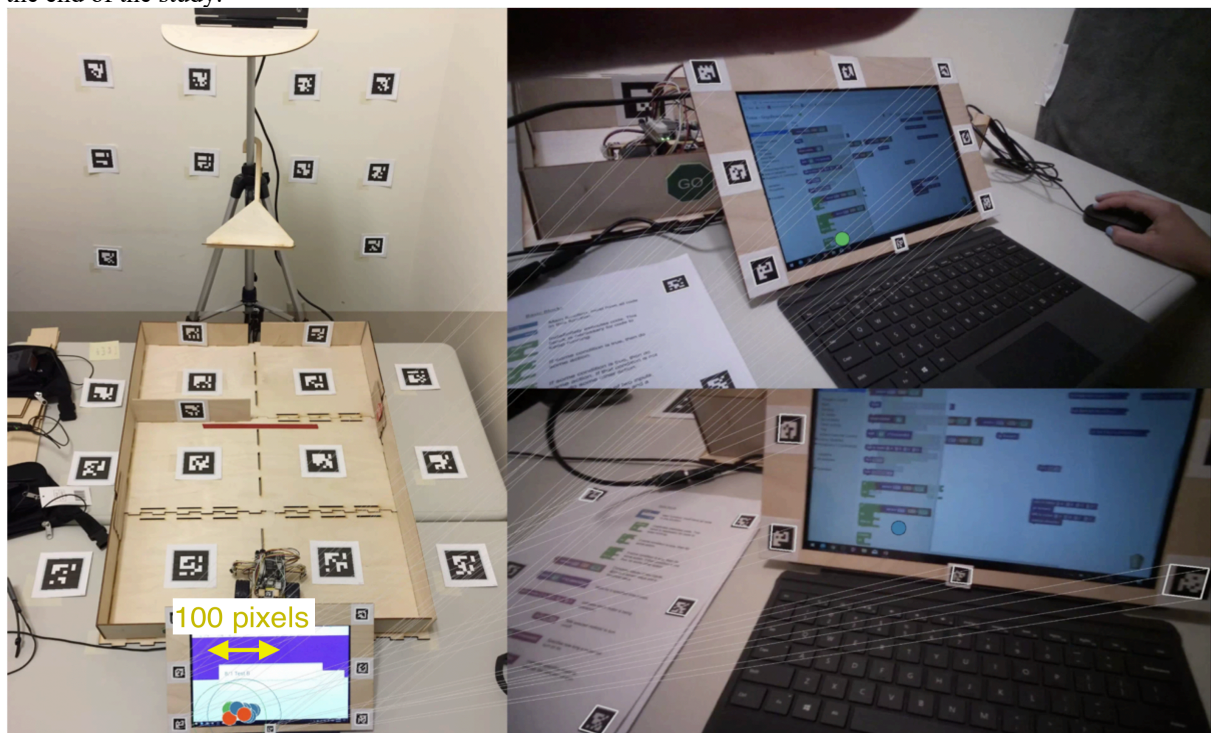
1. JVA is associated with higher quality of collaboration; more specifically, JVA is associated with participants' ability to sustain mutual understanding (Schneider & Pea, 2013).
2. The number of cycles of individual work (no-JVA) and collaborative interactions (JVA) is positively associated with the three outcome measures (collaboration, task performance, learning gains).

In the next section, I describe the data, preprocessing steps and measures.

### Data collection (multimodal sensors)

Several sensors were used to collect data from both participants in each session. Tobii Pro Glasses 2 eye-tracking glasses were worn by each participant to follow eye gaze relative to a set of fiducial markers placed around the study environment. An Empatica E4 wrist sensor tracked participant electrodermal activity, blood volume pulse, and acceleration. Finally, a Kinect motion sensor was used to track the movement and position of the participants in space. This sensor collects approximately 100 variables related to a person's body joints and skeleton (24 different points with columns for x, y, z coordinates), their facial expressions, and their amount of speech. Typically collected at 30 Hz, this results in roughly 5.4 million observations per individual during a 30-minute session.

This paper focuses on the Tobii eye-tracker and the data it generates. The glasses include multiple cameras (two infrared cameras recording eye movements and one scene camera recording the participant's field of view), an accelerometer, gyroscope, microphone, a wearable recording unit running and associated controller software running on Windows. The Tobii eye-Tracker outputs data of multiple kinds including an audio recording of the session, a video recording from the point of view of the user, the x and y coordinate of the user's eye-gaze relative to its point of view. These glasses sampled at 50 Hz, generating roughly 90,000 observations per person during the main 30-minute activity. No participant reported being bothered by the glasses. Anecdotally, a few participants forgot that they were wearing them and attempted to leave the room without removing the glasses at the end of the study.



**Figure 1.** Example of a video frame generated for sanity-checking purposes where a homography was used to remap participants' gaze (shown in blue and green on the right side of the image) onto a ground truth (left side).

The white lines represent the points detected from the fiducial markers to do the homography. On the left, the gaze points turn red if there are within a certain radius (e.g., 100px), which signifies some joint visual attention.

### Data preprocessing – Temporal and spatial alignment

#### Temporal alignment

In order to clearly mark when transitions between different portions of the study took place across all sensors and recording devices, several fiducial markers with accompanying audio cues were placed in a PowerPoint presentation used by researchers and participants to guide the flow of the study. Whenever specific points in the study were reached, participants would simultaneously see the fiducial marker, hear the sound, and press a button on the EDA bracelet. In this way, all of the sensors on both participants as well as the video recorder would have

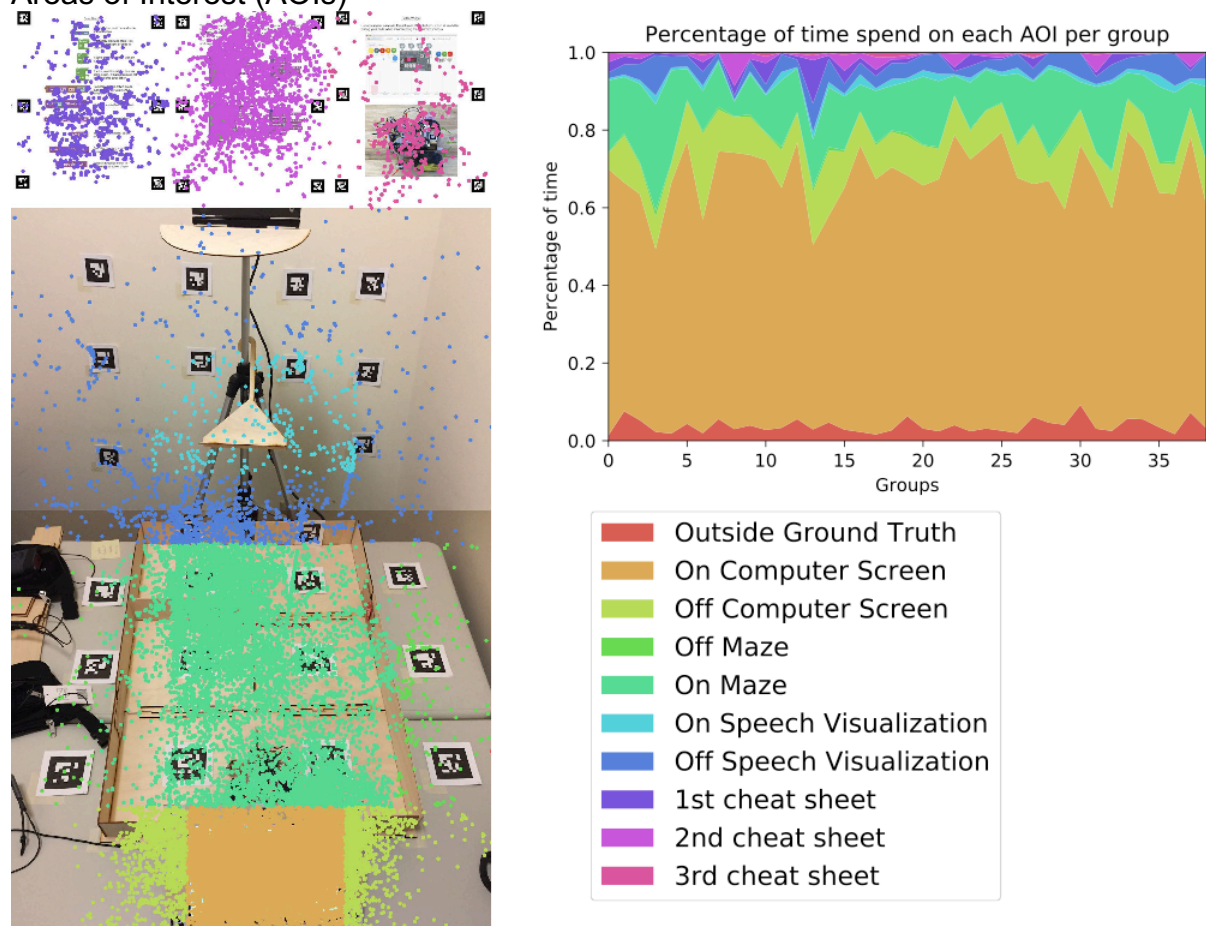
some tagged record of the event and therefore a way to synchronize all of the data. The eye-tracking data analyzed in this paper is solely from the main 30-minute portion of the study and was synchronized via these tags.

### Spatial alignment

One challenge of using mobile eye-trackers is that users can freely move - they can walk around, stand, sit, and change the orientation of their head. Their eye gaze is calibrated on the frames provided by the scene camera, which changes in its content depending on where users are looking. This kind of data is significantly more challenging to analyze compared to traditional (i.e. remote) eye-trackers, where the main area of interest is the screen of a computer. Thus, when using a mobile eye-tracker, we need to identify which part of the environment users are looking at. The solution used in this paper is to add fiducial markers to the environment (they look like QR codes on Figure 1). Detecting those markers is relatively easy for computer vision algorithms, and since they each have a unique ID they provide common coordinates across different perspectives. More specifically, a panoramic picture of the workspace (Fig. 1, left side) and the markers detected from the scene camera of the mobile eye-tracker (Fig. 1, right side) were associated to the markers of the workspace (referred to as “ground truth” below). Knowing this common set of points allowed to infer the location of users’ gaze points on a common plane using a *homography*. The left side of Figure 1 shows the last 5 gaze points for each user (shown as a gaze plot; additionally, the dots turn red if they are within 100 pixels of each other). Finally, for each group a video recording was generated for sanity checking purposes and to make sure that the homography was accurate.

## Results

### Areas of Interest (AOIs)



**Figure 2.** AOIs on the ground truth and Gaze points on each AOI. On the left: the distribution of gaze points for one group (42). On the right: the percentage of eye-tracking data for each AOI (y-axis) and for each group (x-axis). “Outside Ground Truth” refers to the gaze points outside the image shown on the left side of Fig. 1.

Areas of Interests (AOIs) divide the participants view into 7 different regions, at three different height levels. At the lower level, they differentiate between looking at the computer screen where participants wrote code and



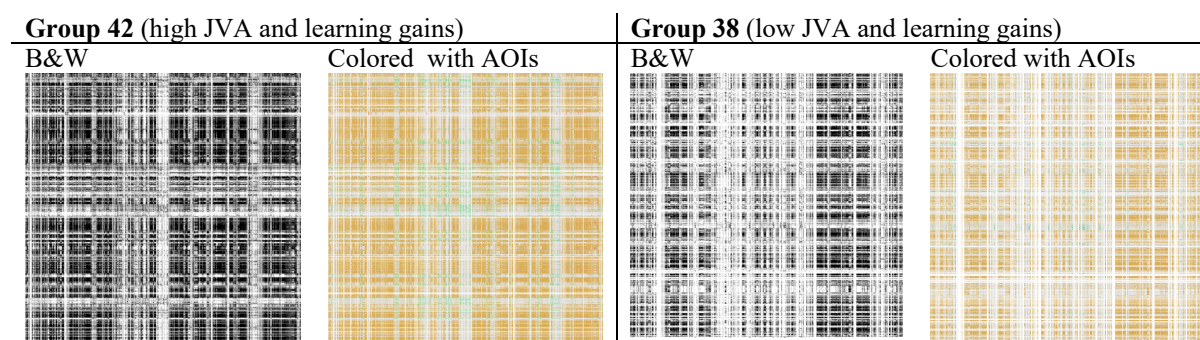
looking around it. At the level of the maze, they differentiate gaze points within the maze and outside the maze. Finally, at the level of the wall, they separate the area corresponding to the speech visualization (only relevant to groups having access to it, i.e. condition #2 and #4) from the one around it.

Running Pearson's correlations between the number of gaze points on each AOI at the individual level generated the following results: looking at the maze and code quality ( $r(37) = 0.331, p = 0.040$ ) / learning gains ( $r(37) = 0.360, p = 0.025$ ). Additionally, there were negative correlations between looking at the computer screen and code quality ( $r(37) = -0.320, p = 0.047$ ); looking at the first cheat sheet and sustaining mutual understanding ( $r(36) = -0.364, p = 0.025$ ) / quality of collaboration ( $r(36) = -0.323, p = 0.048$ ); looking at the second cheat sheet and task performance ( $r(36) = -0.608, p < 0.001$ ) / code quality ( $r(37) = -0.350, p = 0.029$ ). In summary, looking at the number of times that individual participants looked at different AOIs seemed to be mostly associated with negative outcomes.

### Cross-recurrence graphs

Before computing measures of joint visual attention, it is recommended to generate cross-recurrence graphs to sanity check the data. A Cross-Recurrence Graph (Jermann, Mullins, Nüssli & Dillenbourg, 2011) is a plot representing the eye-tracking data of the dyad. One axis is the time for one person and the other axis is the time for the other participant. If the two people are looking at the same location at the same time, we plot a black dot along the diagonal. If there is a delay, we plot this point above and below the diagonal. The distance from the diagonal is proportional to the delay. Therefore, by looking at the points on the diagonal we can estimate visual coupling within the pair. Gray dots represent no joint visual attention and white dots represent missing data.

Cross-recurrence graphs provided a visual representation of the groups' attentional alignment. I generated one for each group and used them as a sanity check for the JVA measures (Fig. 3): for example, it confirmed that group 42 had high levels of JVA, which is represented by more black pixels. Group 38 had low levels of JVA which is represented by more white pixels. Color-coded cross-recurrence graphs (i.e., using the colors from Fig. 3) also helped us observe patterns of interaction: groups spent most of their time looking at the computer screen (gold), and the maze (green).



**Figure 3.** Cross-recurrence graphs for two dyads. The left side shows moments of joint visual attention (black), no joint visual attention (gray) and missing data (white). The right side shows joint attention on particular AOIs (gold = computer screen, green = maze). The two graphs on the left show a productive group with high learning gains (42), and the two graphs on the right shows a group with low learning gains (38).

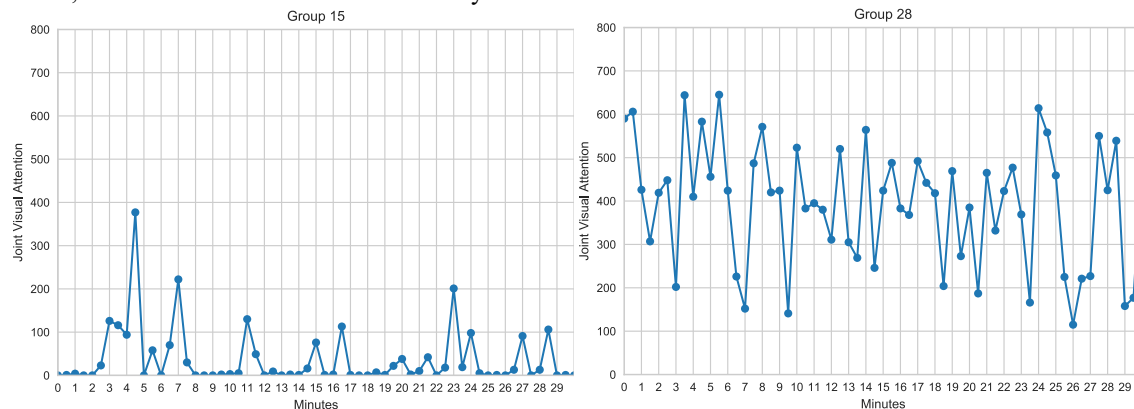
### Joint Visual Attention

Joint Visual Attention (JVA) was computed according to prior research (Richard & Dale, 2015; Schneider & al., 2018; Schneider & Pea, 2013; Gergle & Clark, 2011). As a first pass, I used a radius of 100 pixels for two gaze points to be considered as JVA (this radius is shown on Figure 1). For each gaze point, I also checked whether the other group member looked at the same area within +/- 2 seconds (which has been shown to be the amount of time necessary for someone to disengage from what they are doing and pay attention to a partner's actions; Richardson). The results are summarized below.

I found significant correlations between JVA and: "sustaining mutual understanding":  $r(36) = 0.397, p = 0.014$ ; "task division":  $r(36) = 0.351, p = 0.031$ ; and their overall quality of collaboration:  $r(36) = 0.341, p = 0.036$  (see Meier, Spada, & Rummel, 2007 for a definition of those constructs). There was no significant correlation with task performance or learning gains. It should be noted that I also looked at other radius sizes in addition to 100 pixels (50, 150, and 200 pixels) which defined the distance between two gaze points to be considered as a moment of joint visual attention. In these analyses, the size of the radius did not influence the correlations reported above.

### Cycles of collaboration (JVA) and individual work (no-JVA)

For those analyses, I look at the number of times participants shifted between collaborative work (i.e., with increased levels of joint visual attention) and individual work (i.e., with lower levels of joint visual attention). I tried several approaches and found that the following steps provided the most conclusive measure: 1) I summed the number of moments of JVA for different time windows (a 60 second time window is shown on Fig. 4); 2) I compared each observation with the previous point and looked at whether JVA was going up or down; 3) I counted the number of times the group shifted from increasing to decreasing (and vice versa) their levels of JVA. In other words, this measure offers an estimate for cycles of individual and collaborative work.



**Figure 4.** Levels of joint visual attention over time for two groups. Group 15 had the lowest number of cycles according to the measure above (12) and group 28 had the larger number of cycles (24).

This measure was correlated with learning gains using 30 seconds increment (i.e., 30sec., 60sec., 90sec., 120sec.). I found significant correlations with a 30 second window:  $r(35) = 0.349$ ,  $p = 0.035$ , 90 sec. window:  $r(35) = 0.355$ ,  $p = 0.031$ , 120 sec. window:  $r(35) = 0.515$ ,  $p = 0.001$  but not with a 60 sec. window:  $r(35) = 0.001$ ,  $p = 0.99$ . Finally, by looking at smaller time windows between 10sec and 60sec., there was a time window (40 sec.) that was significantly correlated with all three dependent measures: overall quality of collaboration  $r(34) = 0.347$ ,  $p = 0.038$ , Task Performance  $r(34) = 0.355$ ,  $p = 0.034$  and Learning gains  $r(35) = 0.398$ ,  $p = 0.015$ . There was no significant correlation between the aggregated JVA measure reported in the section above (i.e., the total amount of JVA) and the measure described in this section - which suggests that they are capturing two different constructs. I discuss these results below.

## Discussion

This paper replicates prior results showing that JVA is positively correlated with high quality collaborative interactions (Jermann, Mullins, Nüssli & Dillenbourg, 2011; Schneider & Pea, 2013; Schneider, Sharma, Cuendet, Zufferey, Dillenbourg & Pea, 2018). It also provides further evidence that JVA can be computed in a co-located setting using fiducial markers disseminated in the environment. Since most of the prior work was done in remote settings, it is timely that new approaches allow researchers to study collaborative learning in more ecological ways. The final and main contribution of this paper is a new measure that captures cycles of collaboration and individual work in dyads. This measure provides a complementary lens into collaborative processes: I found JVA to be positively associated with collaboration quality, and this new measure with learning gains (as well as task performance and collaboration quality, depending on the threshold used). This suggests that an important feature of successful collaborative learning groups is to balance individual cognition with group work. While this is beyond the scope of this paper, future work will study this effect in more detail by qualitatively analyzing dyads that are driving this effect (i.e., groups with low learning gains and low scores on this measure, and dyad with high learning gains and high scores on this measure). Additionally, I am planning to replicate the findings above on a different dataset, which would provide further evidence that cycles of collaboration and individual work positively contribute to learning.

## Conclusions

This paper presents a study where dyads of participants worked on programming a robot to solve a variety of mazes. I found that Joint Visual Attention can be captured using dual eye-trackers in co-located settings, and that this measure is positively correlated with collaboration quality. Additionally, I designed a new measure intended to capture cycles of individual work and group collaboration. This measure shed a new light on what constitutes productive interaction in dyads of students.

It should be acknowledged that this paper has several limitations (for a discussion of the limitations related to the task and the dependent measures used, please refer to Starr, Reilly & Schneider, 2018). First, this

paper mostly relied on correlations. While this provides intuitive results, future work should use more comprehensive statistical tests to model participants' interactions and control for collinearity. Second, the new measure presented in this paper relies on several parameters (minimum distance between two gazes to qualify as joint visual attention, different time windows) that were arbitrarily defined. A better understanding of how those parameters need to be fine-tuned is important for generalizing this measure to other settings. Finally, dual eye-tracking only offers a limited view of collaborative processes. Future work will integrate sensor data from multiple modalities (e.g., electrodermal, motion and speech data) to get a more complete picture of what constitutes productive interactions in co-located settings.

In conclusion, this study shows that it is possible to develop new ways of capturing 21st century skills in hands-on tasks typical of makerspaces. Even with the limitations mentioned above, this work makes a first step in this direction and opens the way to more rigorously studying collaborative processes in open-ended learning environments using dual mobile eye-trackers.

## References

- Berland, M., Baker, R. S., & Blikstein, P. (2014). Educational data mining and learning analytics: Applications to constructionist research. *Technology, Knowledge and Learning*, 19(1–2), 205–220.
- Blikstein, P., & Worsley, M. (2016). Multimodal Learning Analytics and Education Data Mining: using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, 3(2), 220–238.
- Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). Washington, DC, US: American Psychological Association.
- D'Angelo, S., & Begel, A. (2017). Improving Communication Between Pair Programmers Using Shared Gaze Awareness. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 6245–6290). New York, NY, USA: ACM.
- Dede, C. (2010). Comparing frameworks for 21st century skills. *21st Century Skills: Rethinking How Students Learn*, 20, 51–76.
- Gergle, D., & Clark, A. T. (2011). See What I'M Saying?: Using Dyadic Mobile Eye Tracking to Study Collaborative Reference. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work* (pp. 435–444). New York, NY, USA: ACM.
- Jermann, P., Mullins, D., Nüssli, M.-A., & Dillenbourg, P. (2011). Collaborative Gaze Footprints: Correlates of Interaction Quality. *Connecting Computer-Supported Collaborative Learning to Policy and Practice: CSCL2011 Conference Proceedings., Volume I-Long Papers*, 184–191.
- Meier, A., Spada, H., & Rummel, N. (2007). A rating scheme for assessing the quality of computer-supported collaboration processes. *International Journal of Computer-Supported Collaborative Learning*, 2(1), 63–86.
- Miyake, N. (1986). Constructive interaction and the iterative process of understanding. *Cognitive Science*, 10 (2), 151–177.
- Richardson, D. C., & Dale, R. (2005a). Looking To Understand: The Coupling Between Speakers' and Listeners' Eye Movements and Its Relationship to Discourse Comprehension. *Cognitive Science*, 29(6), 1045–1060.
- Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007). The Art of Conversation Is Coordination Common Ground and the Coupling of Eye Movements During Dialogue. *Psychological Science*, 18(5), 407–413.
- Schneider, B., & Pea, R. (2013). Real-time mutual gaze perception enhances collaborative learning and collaboration quality. *International Journal of Computer-Supported Collaborative Learning*, 8(4), 375–397.
- Schneider, B., Sharma, K., Cuendet, S., Zufferey, G., Dillenbourg, P., & Pea, R. (2018). Leveraging Mobile Eye-Trackers to Capture Joint Visual Attention in Co-Located Collaborative Learning Groups. *International Journal of Computer-Supported Collaborative Learning*.
- Starr, E., Reilly, J., & Schneider, B. (2018). Toward Using Multi-Modal Learning Analytics to Support and Measure Collaboration in Co-Located Dyads. *ICLS2018 Conference Proceedings., Volume I-Long Papers*, 448–455.
- Tomasello, M. (1995). Joint attention as social cognition. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 103–130). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.
- Tschan, F. (2002). Ideal Cycles of Communication (or Cognitions) in Triads, Dyads, and Individuals. *Small Group Research*, 33(6), 615–643.
- Yu, C., & Smith, L. B. (2013). Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination. *PloS One*, 8(11), e79659.

# The Lifespan and Impact of Students' Ideas Shared During Classroom Science Inquiry

Camillia Matuk, New York University, [cmatuk@nyu.edu](mailto:cmatuk@nyu.edu)  
Wanjing (Anya) Ma, University of Pennsylvania, [wm821@nyu.edu](mailto:wm821@nyu.edu)  
Garima Sharma, New York University, [garimasharma@nyu.edu](mailto:garimasharma@nyu.edu)  
Marcia C. Linn, University of California, Berkeley, [mclinn@berkeley.edu](mailto:mclinn@berkeley.edu)

**Abstract:** Sharing ideas can strengthen students' science explanations. Yet, how to guide uses of peers' ideas, and what the impacts of those ideas are on students' learning, are open questions. We implemented a web-based cell biology unit with 116 grade 7 students, and explored how peers' ideas are used during explanation building, and how prompts to draw on peers to either diversify or reinforce existing ideas impacted the quality of students' written explanations. Among other findings, exchanging ideas with peers led to all students improving their explanation quality upon revision; and students prompted to diversify their ideas showed greater learning gains by the end of the unit, while students prompted to reinforce ideas, who used more peer-generated ideas in preparation to write their explanations, produced higher quality explanations. This study builds our understanding of the influence of peer ideas on learning, and offers insight into supporting students in engaging effectively with peers' ideas.

## The role of peer ideas in knowledge integration

Because scientific knowledge is socially constructed (Latour & Woolgar, 1979; Lemke, 1990), the ability for science learners to engage productively with one another's ideas is critical to develop, as is their ability to refine explanations and arguments in light of new ideas (NGSS Lead States, 2013; Kuhn, 2012). Nurturing a culture around sharing and improving collective understanding can position students as creators, rather than just readers of knowledge (Scardamalia & Bereiter, 2014). However, research is mixed on the best ways to support students' encounters with others' ideas (De Jonge, 2018). For example, one study documented how the ideas that middle school students collected during their inquiry investigations related to the quality of their later scientific explanations. Specifically, students' tendencies to find ideas that reinforce, as opposed to diversify their existing repertoire of ideas before writing an explanation, and their tendencies to self-generate those ideas rather than to use peer-generated ideas, also resulted in them writing stronger scientific explanations (Matuk & Linn, 2018). Another study suggests that diversifying rather than reinforcing ideas is a more successful strategy for students of higher prior knowledge than it is for students of lower prior knowledge (Matuk & Linn, under review). These findings begin to suggest how different ways of engaging with peers' ideas can support students' own thinking; and for whom these ways may be more or less productive.

Our research is guided by the Knowledge Integration (KI) framework (Linn & Eylon, 2011), a perspective based in research in the learning sciences, and that views learners as constructing understanding by distinguishing among various new and existing ideas. KI instruction supports learning by eliciting students' existing ideas, assisting them in distinguishing these from newly encountered ideas, and organizing these into a coherent understanding. However, the role of peers in knowledge integration is less clear. Little work has focused, for example, on the lifespan of shared ideas in a classroom. That is, when students share ideas, which types of ideas are taken up and used throughout the course of students' inquiry? How do these shared ideas support students in constructing sound explanations? Moreover, what strategies for using these ideas might be more or less effective?

This study extends prior research by exploring the kinds of ideas that students exchange during inquiry, and ultimately incorporate into their explanations. It also explores the value of different strategies for using peers' ideas on the quality of their explanations, and on their overall learning outcomes.

## Research questions

We implemented a web-based science inquiry unit in a middle school classroom, which features a tool that supports students in sharing ideas with classmates as they work toward constructing scientific explanations. Our specific research questions are:

1. What overall impact does the unit have on students' conceptual learning? How does prompting students to seek similar vs. different ideas from peers impact their overall learning?
2. How does prompting students to seek similar vs. different ideas from peers impact the quality of



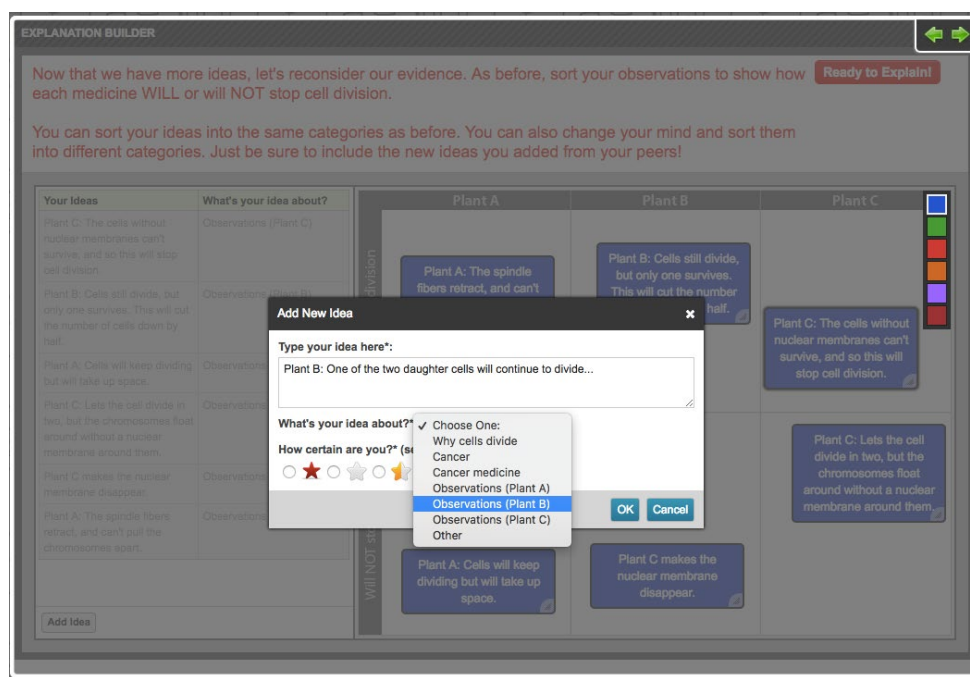
students' explanations?

3. What is the trajectory of ideas across students' explanation building process? That is, how did students take up and incorporate their peers' shared ideas? Which kinds of ideas were more popular? How did peers' ideas impact the quality of students' explanations?

## Methods

### WISE and the Idea Manager

The Web-based Inquiry Science Environment, or WISE ([wise.berkeley.edu](http://wise.berkeley.edu)) is a free, open-source platform created to design and deliver classroom-based science inquiry curriculum (Slotta & Linn, 2009). The Idea Manager is a tool integrated into WISE to support students in collecting, distinguishing, and organizing their ideas into coherent science explanations during their inquiry investigations (Matuk et al., 2016). Following the KI framework, the Idea Manager offers a persistent space called the *Private Idea Basket*, within which students can document their existing and new ideas in brief entries; sort and distinguish among these ideas using a visual organizer called the *Explanation Builder*; and refer to their organized ideas to integrate them into a written explanation (Figure 1). Students can choose to add any of their idea entries to a *Public Basket*, which anonymously lists all ideas shared by students in the same class. Students can select publicly shared ideas to “copy” into their own Private Baskets. Thus, peer-generated ideas become available for use alongside students' self-generated ideas.



**Figure 1.** Foreground: An idea entry dialog box. Background: the Explanation Builder, in which students organize ideas, then refer to them to write an explanation (by clicking the “Ready to Explain” button).

### The *Mitosis* unit

The unit, *What makes a good cancer medicine?: Observing mitosis and cell processes (Mitosis)*, introduced students to the process of cell division and its relationship to cancer and cancer treatment. Animations, diagrams, and narrative explanations introduced students to the phases of normal cell division, and to the notion that cancer is a disease in which cells have lost the mechanism that controls their division. An effective cancer treatment must thus stop cancerous cells from dividing, but this does not occur without side effects.

Students then compared the effectiveness of three potential, plant-derived cancer medicines. They viewed animations of dividing cells that were treated by each medicine, and noted their observations in their Private Idea Baskets. Following this, they organized their ideas in the Explanation Builder to sort pros and cons for each medicine, and then referred to these ideas to write an explanation for the medicine they recommended.

Students were then asked to make public the ideas they used in their recommendations, and to select public ideas from their peers to add to their own Idea Baskets. They then reorganized their previously and newly

added ideas within the Explanation Builder and referred to these to revise their recommendations (Figure 2).

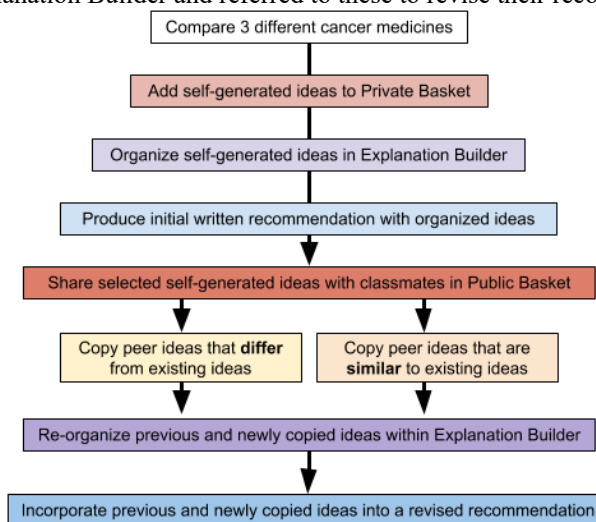


Figure 2. The *Mitosis* activity sequence and study design.

## Participants and study design

Participants were 144 grade 7 students across 5 class periods of one teacher, from a diverse public middle school on the West coast of the United States. Students worked on the unit in pairs at their own pace during class time for 10 consecutive school days. The teacher, who had 5+ years of experience teaching with WISE, circulated to assist students, led occasional whole class discussions to address common conceptual challenges, and offered guidance on upcoming unit activities. Collaboration occurred at different levels: First, by working on shared computers, student partners had to come to consensus through discussion over their responses. Second, the teacher regularly reminded students to document and share their ideas with others in class, such as by pointing out good ideas during her conversations with partners, and encouraging them to add these to the Public Basket.

Students individually completed a pre and posttest on the first and last days of the study. One of the three items on this pre/posttest addressed students' understanding of the order and importance of the phases of cell division. A second item tested their ability to apply this understanding to reason about the action of an effective cancer medicine. A third item asked students to select ideas from fictional peers that would help them to write an explanation for the role of spindle fibers during cell division.

To investigate the value of different strategies for using peers' ideas, students were divided into two conditions. In the Reinforce condition (N=66 students, 36 workgroups), students were prompted to collect peer ideas that were similar to their own ideas. In the Diversify condition (N=50 students, 27 workgroups), they were prompted to collect peer ideas that differed from their own. We excluded from our analysis students who had not completed the pre/posttest, nor submitted both initial and revised recommendations. The total number of students in our dataset was therefore 116.

## Data and analysis

Data consisted of each workgroup's individual ideas, including information on which ideas they kept private and which they made public, which public ideas were copied by which workgroups, and which ideas were used in students' Explanation Builders. We also collected students' initial and revised written recommendations and their responses to the pre/posttest.

To determine differences between conditions in the overall impact of the unit on students' learning outcomes (RQ1), we scored students' pre and posttests based on previously developed KI rubrics, which give credit to responses that integrate key ideas (Matuk & Linn, 2018). Two independent raters achieved high inter-rater reliability across the three items ( $\kappa=0.91, 0.83$  and  $0.71$ ). We summed scores across items to obtain a total score for each student, and used a t-test to detect differences in mean total scores between conditions.

To compare the effects on students' recommendations of prompts to seek peer ideas that differed and that resembled their own (RQ2), we scored students' initial and revised recommendations. One rubric identified the presence of *Key Concepts* (Table 1) and another rated the *Argument Structure* (Table 2), while the sum of the scores on these two criteria produced an overall score out of 10 of the explanation's quality. Two independent raters scored 20–30 student responses at a time, and resolved disagreements through discussion

until inter-rater reliability had been achieved (*Key Concepts*  $\kappa=0.83$ ; *Argument Structure*  $\kappa=0.96$ ). We used t-tests to detect differences in students' gains on the individual *Key Concepts* and *Argument Structure* dimensions, and on their overall explanation quality, between workgroups' initial and revised recommendations, and between the Reinforce and Diversify conditions.

Table 1: Rubric for identifying *Key Concepts* in students' initial and revised recommendations

| Key Concepts  |   |  |
|---|---|--|
| <ul style="list-style-type: none"> <li>• <i>Org</i>: Identifies a cell organelle to be affected by the medicine (e.g., chromosomes, spindle fibers).</li> <li>• <i>Process</i>: Explains the manner by which the medicine will disrupt the organelle's function (e.g., will prevent chromosomes from dividing).</li> <li>• <i>Need</i>: Mentions that the medicine must stop cell division/cancer growth in order to treat cancer.</li> </ul> |   |  |
| Score   | Criteria  | Examples   |
| 0/1   | No Answer/Offtask or uninterpretable                    | (Blank)/I don't know.  |
| 2   | Irrelevant, incorrect or ambiguous                      | We think plant A will work the best because it kills off the cell without leaving with empty cells that clog the body which is what cancer does as well.   |
| 3   | Any ONE of the three key ideas is correctly explained.  | I think that the best cancer medicine is plant A, because it will <u>stop the cell from undergoing mitosis</u> [ <i>Need</i> ]. It will stop the the [sic] cell from ever doing mitosis, but the cell might still be able to survive.  |
| 4   | Any TWO of the three key ideas are correctly explained. | We will recommend plant A because it stopped the <u>spindle fibers</u> [ <i>Org</i> ] from working before a new cell was created therefore it <u>stopped mitosis</u> [ <i>Need</i> ].  |
| 5   | All THREE key ideas are correctly explained.            | Plant A because it is the most effective at <u>stopping mitosis</u> [ <i>Need</i> ]. The pros to this plant is that the <u>spindle fibers</u> [ <i>Org</i> ] don't go all the way so mitosis isn't complete. So if the spindle fibers <u>don't go all the way</u> [ <i>Fxn/Process</i> ] through then the <u>chromosomes</u> [ <i>Org</i> ] won't <u>line up</u> [ <i>Process</i> ]. The cons are that their will be side effects from the medicine. |

Table 2: Rubric for evaluating the *Argument Structure* of students' initial and revised recommendations

| Score | Criteria   | Examples  |
|-------|--|---|
| 0/1   | No Answer/Offtask or uninterpretable   | (Blank)/I don't know.   |
| 2     | Claims stated with no supporting evidence.   | I think Plant A is most effective.  |
| 3     | Interpretation given that includes a claim supported by evidence, which may or may not properly align. No counter-arguments are given.   | we think that the most effective cancer medicine is plant A due to the fact that it stopped the spindle fibers from growing and the spindle fibers will not grow to redirect the chromosomes [ <i>supported interpretation</i> ]  |
| 4     | Interpretation given that includes a claim aligned with supporting evidence. Offers unelaborated counter-arguments (e.g. pros/cons, side effects, comparison to alternatives). | Plant A because it is the most effective at stopping mitosis [ <i>supported interpretation</i> ]. The pros to this plant is that the spindle fibers don't go all the way so mitosis isn't complete. So if the spindle fibers don't go all the way through then the chromosomes won't line up. The cons are that their will be side effects from the medicine [ <i>unelaborated counterargument</i> ].             |
| 5     | Interpretation given that includes a claim aligned with supporting evidence. Offers at least one elaborated counter-argument (e.g. pros/cons, side effects, or comparison).    | We think that medicine A is the best because it stops the cell from dividing, which would work the best [ <i>valid interpretation</i> ]. Medicine B only destroys the second pair of chromosomes, which means when the medicine wears off, the cell can split again. Medicine C only destroys the cells' membranes which means that the cells might still be able to split [ <i>elaborated counterargument</i> ]. |

To track the trajectory of ideas (RQ3), we gave each idea a unique “private idea ID” that indicated the workgroup that had generated it; and each copied idea a unique “copied idea ID” that indicated the workgroup that copied it. These IDs allowed us to identify the trajectory of each idea from its origin, to its sharing in the Public Basket, and to its use in an Explanation Builder step. Finally, we used these IDs to note which ideas were

present in students' written recommendations. We used t-tests to detect significant differences in students' engagement with ideas across the unit. We also tested how the proportion of peer- vs. self-generated ideas impacted the quality of students' revised recommendations.

Because students were collecting ideas to construct arguments for the medicines they recommended, we characterized the ideas in terms of three main components of an argument: *Claim*, *Observation*, and *Interpretation* (Table 3). These categories are based on theories of argument structure, which describe the function of argument to be in creating links between claims and evidence (Burlison, 1979; Kneupper, 1979; Toulmin, 2003). Two independent raters scored 20 ideas at a time until they had achieved Cohen's kappa values of 0.83, 0.90, and 0.83 on *Claim*, *Observation*, and *Interpretation*, respectively. One coder then categorized the rest of the ideas. We excluded from our analyses nine of the ideas that were irrelevant to the task of constructing a recommendation (e.g., "Three plant idea").

We determined the relative *popularity* of ideas by the number of times these were copied by other workgroups. We then used an ANOVA to uncover associations between the kinds of ideas and their popularity.

**Table 3: Scoring rubrics for distinguishing ideas as Claim, Observation, and Interpretation**

| Type           | Description   | Example  |
|----------------|---|--|
| Claim          | Declarative, factual statements not supported by evidence from the unit.                                      | Cancer is the overgrowth of cells. It can travel down the body in the blood stream.              |
| Observation    | Descriptions, without accompanying inferences, of information from the unit (e.g., of animations).            | The treated cell's spindle fibers stopped. They didn't grab the chromosomes. They're also green. |
| Interpretation | Causal/explanatory statements or inferences that integrate ideas that are not otherwise explicit in the unit. | The mitosis is not prevented but slowed down because the daughter cells wont be able to divide.  |

## Findings

### RQ1: How did the unit impact students' learning overall?

The unit as a whole appeared to positively impact students' conceptual learning. All students showed significant gains between the pre and posttest (N=116; M=3.16; SD=2.27;  $t(115)=15.00$ ,  $p<.001$ ,  $d=1.63$ ). Prompts for students to use peers' ideas to diversify their own ideas seemed to be particularly beneficial, as students in the Diversify condition (N=50; M=3.74; SD=2.26) showed significantly greater gains than students in the Reinforce condition (N=66, M=2.73, SD=2.20),  $t(114)=-2.43$ ,  $p=.017$ ,  $d=.45$ ).

### RQ2: How did exchanging ideas impact students' written explanations?

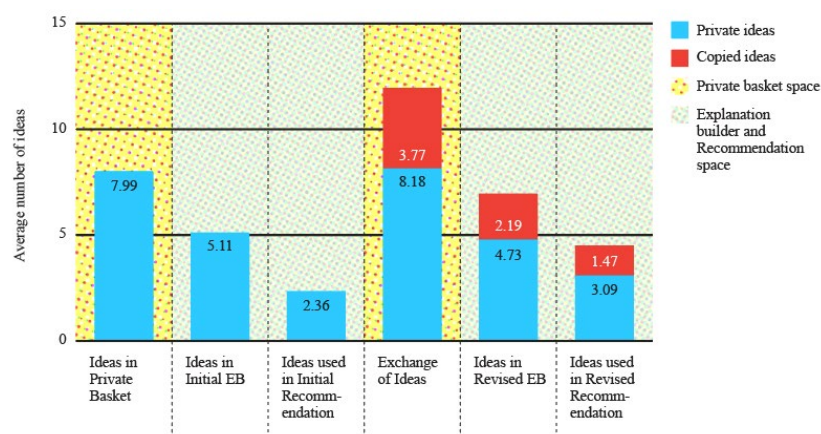
Students' overall recommendations improved significantly upon revision following their exchange of ideas with their peers (M=1.14; SD=1.80;  $t(62)=5.05$ ,  $p<.001$ ,  $d=.75$ ). In both conditions, students' final recommendations reflected more key concepts (M=0.57, SD=1.04,  $t(62)=4.35$ ,  $p<.001$ ,  $d=.58$ ) and better argument structure (M=0.57; SD=0.14;  $t(62)=4.11$ ,  $p<.001$ ,  $d=.65$ ) than their initial recommendations. There was no significant difference in the overall explanation quality between the Reinforce condition (M=8.94, SD=1.59, N=36) and the Diversify condition (M=8.78, SD=1.37, N=27),  $t(61)=0.438$ ,  $p=.663$ ,  $d=.108$ . There were also no significant differences in gains between the Key Concepts and Argument Structure criteria within nor between conditions. These findings suggest that exchanging ideas generally had a positive impact on students' revisions, regardless of the manner by which students were prompted use their peers' ideas.

### RQ3: What was the trajectory of ideas across the unit?

#### How did students take up and incorporate their peers' ideas into their explanations?

To understand the overall trajectory of self- and peer-generated ideas, we graphed the average number of ideas at six time points (Figure 3). These included the ideas before and leading up to the initial recommendation, specifically, the ideas (1) generated in students' Private Baskets (Reinforce, M=7.33; Diversify, M=8.66); (2) organized in the Explanation Builder (Reinforce, M=5.02; Diversify, M=5.22); and (3) incorporated into the initial written recommendations (Reinforce, M=2.22; Diversify, M=2.55). Following students' exchange of ideas with their peers, we tracked (4) the copied ideas that students added to their Private Baskets (Reinforce, M=3.86; Diversify, M=3.33); as well as the additional private ideas generated at this point (Reinforce, M=7.58; Diversify, M=8.99); (5) the private ideas re-organized in their Explanation Builders (Reinforce, M=4.61; Diversify, M=4.88); and the copied ideas organized alongside them (Reinforce, M=2.08; Diversify, M=2.33).

Finally, we noted (6) which self-generated ideas (Reinforce,  $M=3.16$ ; Diversify,  $M=3$ ), and which peer-generated ideas students incorporated into their revised recommendations (Reinforce,  $M=1.69$ ; Diversify,  $M=1.18$ ).



**Figure 3.** Trajectory of ideas generated by students ( $N=63$  workgroups) across both conditions. (EB=Explanation Builder)

Figure 3 shows that across conditions, students generated more ideas ( $M=7.9$ ) than they organized in their Explanation Builders ( $M=5.11$ ), and used in their initial recommendations ( $M=2.38$ ). This same pattern is seen in students' uses of their peers' ideas. Specifically, students copied more of their peers' ideas ( $M=3.77$ ) than they organized in their Explanation Builders ( $M=2.19$ ), and incorporated into their revised recommendations ( $M=1.47$ ). Notably, students also generated more ( $M=8.18$ ), and incorporated more of their own ideas, as opposed to their peers' ideas, into their revised recommendations ( $M=3.09$ ) following their exchange with their peers.

Compared to students' uses of ideas leading up to their initial recommendations, students added more ideas to their Private Baskets ( $N=63$ ;  $M=4.06$ ;  $SD=3.33$ ;  $t(62)=6.70$ ,  $p<.001$ ,  $d=.88$ ), organized more ideas in their Explanation Builders ( $N=63$ ;  $M=1.81$ ;  $SD=1.87$ ;  $t(62)=7.70$ ,  $p<.001$ ,  $d=.72$ ), and incorporated more ideas into their final recommendations ( $N=63$ ;  $M=2.21$ ;  $SD=0.25$ ;  $t(62)=8.68$ ;  $p<.001$ ,  $d=1.01$ ) following their exchange of ideas with peers. Students also used significantly more self-generated than peer-generated ideas in their revised Explanation Builders ( $N=63$ ;  $M=2.54$ ;  $SD=2.75$ ;  $t(62)=7.43$ ,  $p<.001$ ,  $d=1.29$ ) as well as in their revised recommendations ( $N=63$ ;  $M=1.62$ ;  $SD=2.22$ ;  $t(62)=5.80$ ,  $p<.001$ ,  $d=0.99$ ) compared to the steps leading up to their initial recommendations.

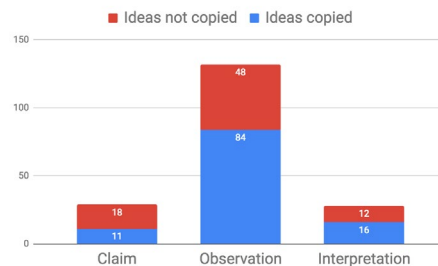
These patterns suggest that students were discerning in the ideas they chose to incorporate into their explanations, a finding that supports the knowledge integration process of distinguishing and sorting among ideas to find the most relevant ones. They also suggest that students used their peers' ideas to complement, rather than to supplant, their own ideas, which attests to the success of the Idea Manager in enabling students to draw upon their peers as supports, rather than being reliant upon them in building explanations.

### Which ideas were most popular?

Across both conditions, *Observations* was the most frequently generated kind of idea (73% Reinforce; 63% Diversify), followed by *Interpretations* (15% Reinforce; 28% Diversify) then *Claims* (12% Reinforce; 18% Diversify). *Observations* was also the most highly copied kind of idea (Reinforce,  $M=2.08$ ; Diversify,  $M=1.93$ ), and significantly more so than either *Interpretations* (Reinforce,  $M=2.00$ ; Diversify,  $M=1.18$ ) or *Claims* (Reinforce,  $M=0.50$ ; Diversify,  $M=0.59$ ), ( $F(2,186)=3.99$ ;  $p=.020$ ) (Figure 4).

It may be that students found observations to be the most persuasive, and *Claims* to be the least robust kinds of peer ideas. For example, the most popular *Claim* across our dataset, copied just 4 times, was "there would be only half as many cells." This idea, which refers to the effect of one of the three medicines on cell division, offers neither supporting evidence, nor an explanation of the consequences of having "half as many cells." Students may thus have regarded it as being of little use for strengthening their own explanations. In contrast, students may have viewed Interpretations, particularly from their peers, to not be as trustworthy as a teacher's explanation, for example. For instance, the most popular Interpretation, copied 8 times, was "Plant A is good because it stops mitosis before it reaches anaphase." This idea offers an explanation, albeit a limited one, for why Plant A is the favored medicine. However, it also offers little detailed evidence, such that students

might be required to trust it at face value. Meanwhile, students may have found observations to be most easily grasped because they could verify these themselves by examining evidence in the unit. For example, the most popular Observation, copied 21 times, was “the plant chemicals don’t stop the division of the cell. the chemicals get rid of the copy of the chromosome so the daughter cell has no chromosomes.” This idea articulates observable events from the animations, which may either offer students new language to rearticulate their existing observations, or highlight details that they might have initially overlooked.



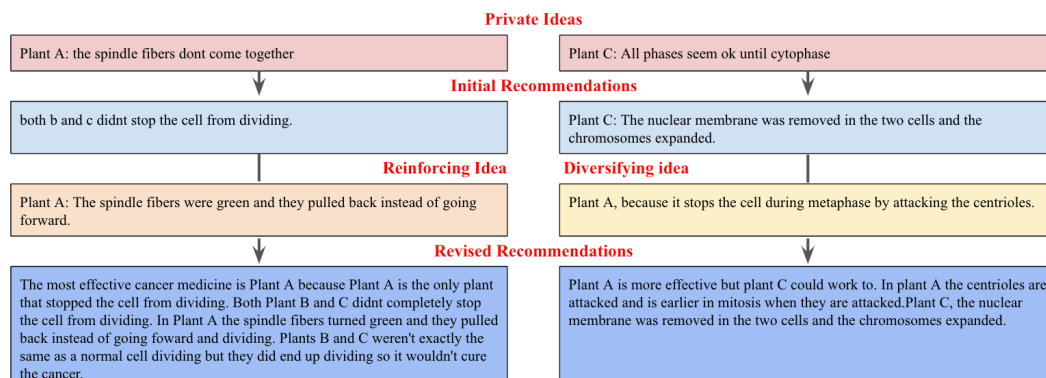
**Figure 4.** Totals of each kind of idea shared within the Public Idea Manager, and the proportion of ideas that were copied vs. not copied by students across both conditions.

**How did peers’ ideas impact the quality of students’ explanations?**

Students in the Reinforce condition who organized a greater proportion of peer- vs. self-generated ideas in the Explanation Builder tended to produce higher quality recommendations, that is, recommendations with scores of 9 or more out of 10,  $t(35)=-2.18, p=.036, d=.78$ . For students in the Diversify condition, however, the relative proportion of peer- vs. self-generated ideas did not have a significant impact on the quality of their explanations ( $t(26)=1.00, p=.327, d=.38$ ). This finding suggests a benefit of organizing peer ideas that reinforced students’ existing ideas.

However, there was no relationship between the proportion of self-generated and peer-generated ideas used in students’ recommendations and the quality of those recommendations. This suggests that students’ improved recommendations were not simply because of access to their peers’ ideas, but more likely because of the effort they put into integrating their own and their peers’ ideas into their explanations.

Figure 5 shows two examples of how students in each condition used their peers’ ideas to improve their recommendations. In both cases, peers’ ideas helped to elaborate recommendations with evidence and to strengthen arguments with counter-arguments.



**Figure 5.** Examples of students’ self-generated ideas, ideas chosen from peers, and initial and revised recommendations. (Left: Workgroup 119924; Right: Workgroup 119945)

**Discussion and significance**

Whereas the unit benefited all students’ learning overall, students prompted to diversify their ideas gained significantly more. Meanwhile, students prompted to reinforce ideas, and who relied more so on peer-generated ideas than on self-generated ideas, produced higher quality in-unit explanations. These findings resonate with prior mixed research on the relative value of diversifying vs. reinforcing ideas. For example, encountering diverse ideas can improve conceptual learning (e.g., Asterhan & Schwarz, 2009; Matuk & Linn, 2018). At the same time, encountering one’s own ideas rearticulated by others can prompt students to revise and improve their own ideas (e.g., Matuk & Linn, 2008; Edge, 2006; Hayes, 2004). Other work suggests that the benefits of reinforcing and diversifying ideas may differ for students with high vs. low prior knowledge (e.g., Matuk &



Linn, under review). Findings from this study suggest that in addition to different students benefiting from different prompts, students may also benefit at different times while building explanations.

Continued analyses might investigate whether there were pre-existing differences between conditions (e.g., prior knowledge) that would offer alternative explanations for our findings, and also whether students actually followed prompts to choose either diversifying or reinforcing ideas. Further research might explore how different students (e.g., high vs. low prior knowledge) benefit from different prompts to use their peers' ideas; and how various strategies (e.g., collecting mostly interpretations vs. observations) impact their success in explanation building.

Visualizing the trajectory of peers' ideas throughout the unit offered insights into the kinds of ideas that students were more or less likely to copy. Importantly, prompts to diversify or to reinforce ideas had no apparent effect on the ways that students used ideas throughout their inquiry. Continued analyses will investigate how the kinds of ideas generated, organized, and used, impact the quality of students' explanations. Future research might also explore other ways to characterize ideas besides in terms of the components of the argument. For instance, we might explore which science concepts are more or less challenging for students to identify at different stages of explanation (cf. McElhaney et al., 2012). Finally, we might examine the role of the teacher in conjunction with technology, in supporting students' productive interactions with peers.

In all, exchanging ideas with peers appeared to positively impact both students' written explanations and their overall learning outcomes. This study adds nuance to the literature on learning from peers. It moreover has implications for technology-rich knowledge sharing supports in classroom-based science inquiry.

## References

- Asterhan, C.S.C. & Schwarz, B.B. (2009). Argumentation and explanation in conceptual change: Indications from protocol analyses of peer-to-peer dialog. *Cognitive Science*, 33, 374-400.
- Burleson, B. R. (1979). On the foundations of rationality: Toulmin, Habermas, and the a priori of reason. *The Journal of the American Forensic Association*, 16(2), 112-127.
- De Jonge, K. M., Rietzschel, E. F., & Van Yperen, N. W. (2018). Stimulated by novelty? The role of psychological needs and perceived creativity. *Personality and Social Psychology Bulletin*.
- Edge, J. (2006). Computer-mediated cooperative development: Non-judgemental discourse in online environments. *Language Teaching Research*, 10(2), 205-227.
- Hayes J.R. (2004) What Triggers Revision?. In: Allal L., Chanquoy L., Largy P. (eds), *Revision Cognitive and Instructional Processes. Studies in Writing*, Vol 13. Springer, Dordrecht
- Kneupper, C. W. (1979). Paradigms and problems: Alternative constructivist/interactionist implications for argumentation theory. *The Journal of the American Forensic Association*, 15(4), 220-227.
- Latour, B., & Woolgar, S. (1979). *Laboratory life: The social construction of scientific facts*. Beverly Hills: Sage.
- Lemke, J. L. (1990). *Talking science: Language, learning, and values*. Norwood, NJ: Ablex Publishing Corporation.
- Linn, M. C., & Eylon, B. S. (2011). *Science learning and instruction: Taking advantage of technology to promote knowledge integration*. New York: Routledge.
- Matuk, C. & Linn, M. C. (2018). Why and how do middle school students exchange ideas during science inquiry? *International Journal of Computer-Supported Collaborative Learning*, 13(3), 263-299.
- Matuk, C., McElhaney, K., King Chen, J., Lim-Breitbart, Kirkpatrick, D. & Linn, M. C. (2016). Iteratively refining a science explanation tool through classroom implementation and stakeholder partnerships. *International Journal of Designs for Learning*, 7(2), 93-110. DOI: 10.14434/ijdl.v7i2.20203
- McElhaney, K., Miller, D., Matuk, C., & Linn, M. C. (2012). Using the Idea Manager to promote coherent understanding of inquiry investigations. In *ICLS'12: Proceedings of the 10<sup>th</sup> International Conference for the Learning Sciences*, (Vol. 1, pp. 323-330). Sydney: Intl Society of the Learning Sciences.
- Scardamalia, M., & Bereiter, C. (2014). Knowledge building and knowledge creation: Theory, pedagogy, and technology. *Cambridge Handbook of the Learning Sciences*, pp. 397-417. Cambridge: Cambridge University Press.
- Slotta, J. D., & Linn, M. C. (2009). *WISE science: Web-based inquiry in the classroom*. New York: Teachers College Press.
- Toulmin, S. E. (2003). *The uses of argument*. Cambridge University Press.

## Acknowledgments

We thank the students and teachers who participated in our study. We also acknowledge funding from the National Science Foundation (Award No. 1119670) and from New York University.

# Collective Knowledge Advancement through Shared Epistemic Agency: Socio-Semantic Network Analyses

Jun Oshima, RECLS Shizuoka Univ., joshima@inf.shizuoka.ac.jp  
Ritsuko Oshima, RECLS Shizuoka Univ., roshima@inf.shizuoka.ac.jp  
Ayano Ohsaki, AIIT, ohsaki-ayano@aiit.ac.jp  
Jin Michael Splichal, Univ. of Helsinki, jin.splichal@helsinki.fi

**Abstract:** This study examined how high-school students engage in collective knowledge advancement through shared epistemic agency during jigsaw instruction, and how their collective knowledge advancement (through shared epistemic agency) relates to learning outcomes. To achieve this, we applied a double-layered socio-semantic network analysis (SSNA). In the first layer of the analysis, we conducted an SSNA to numerically represent collective knowledge advancement and to compare group performances between different levels of learning outcomes. In the second layer, we carried out an Epistemic Network Analysis to examine the relationship between students' shared epistemic agency and learning outcomes. The results revealed that high learning-outcome groups were engaged in epistemic actions to generate new ideas; this led them to quick and sustainable knowledge advancement.

## Research purpose

This study has analyzed the way in which high-school students engage in collective knowledge advancement through shared epistemic agency (Damşa et al., 2010) during jigsaw instruction (Brown & Campione, 1996). It has also investigated the extent to which this collective knowledge advancement relates to learning outcomes. Although studies (e.g., Miyake & Kirschner, 2013) have demonstrated that jigsaw instruction is effective in facilitating conceptual understanding, few studies have shown how learners engage in collective knowledge advancement during collaboration (Oshima, Oshima & Matsuzawa, 2012). The present study explored the students' collective knowledge advancement by applying a double-layered socio-semantic network analysis (SSNA). In the first layer of the analysis, we conducted an SSNA to numerically and visually represent collective knowledge advancement and to compare group performance based on learning outcomes, as evaluated by the SBF framework (in pre- and post-testing). In the second layer, we carried out an Epistemic Network Analysis (ENA) (Shaffer, 2017) to examine the relationship between epistemic frames of shared epistemic agency and learning outcomes.

## Theoretical background

### Inquiry into collective knowledge advancement during jigsaw instruction

Many studies have attempted to design learning environments to facilitate students' collective knowledge advancement. One such instructional method is jigsaw instruction. In jigsaw instruction, learners engage in two different phases of an activity. In the phase known as the "expert group activity," learners work collaboratively on the same materials. Collaboration in the expert group facilitates constructive interaction among the learners (Miyake, 1986), enabling them to develop their own understanding by mutually monitoring each other's ideas from different points of view. After learning their materials during the expert group activity, learners join the jigsaw group activity, where those who studied different materials gather to integrate different sources of knowledge collaboratively. In the jigsaw group activity, learners are predicted to engage in productive interaction within multiple zones of proximal development, depending on the different knowledge sources (e.g., Brown & Campione, 1996). With respect to any one component of the learning activity, one student is an expert on that domain and teaches the other group members. This teaching-learning process is repeated multiple times.

Although studies have demonstrated that jigsaw instruction is effective in facilitating conceptual understanding and collective knowledge advancement (e.g., Miyake & Kirschner, 2014), few studies have examined in detail the process through which learners take responsibility for contributing to collective knowledge advancement through collaboration, i.e., shared epistemic agency. In relation to the knowledge-creation metaphor of learning (Paavola & Hakkarainen, 2005), students are expected to practice knowledge creation through collaboratively constructing knowledge objects (Bereiter, 2002). Scardamalia (2002) has discussed intentional engagement in collective knowledge advancement as epistemic agency—proposing this agency as a new goal for instruction (Scardamalia et al., 2012). Damşa et al. (2010) have also argued in favor of shared epistemic agency, focusing more on group-level agency. Through an in-depth discourse analysis, Damşa et al. (2010) found that



students in collaborative groups engage in the wholly joint epistemic actions of (1) being aware of their lack of knowledge, (2) alleviating the lack of knowledge, (3) creating a shared understanding, and (4) generative collaboration. To regulate their joint epistemic actions, students have also been found to engage in (1) projection by setting goals and creating joint plans, (2) regulation by monitoring and reflecting on their advancement, and (3) developing relations by transcending conflicts, redirecting critical feedback, and creating space for others' contributions. Studies by Scardamalia (2002) and Damşa et al. (2010) suggest that we need to analyze collaboration as knowledge creation, based on at least two layers of collective knowledge advancement: (1) how learners' ideas are improved through their collaborative discourse (i.e., idea improvement) and (2) how learners engage in improving their ideas collectively (i.e., shared epistemic agency).

### Double-layered socio-semantic network analysis

In this study, we propose using two layers of socio-semantic network analyses (SSNA). Recent studies (Oshima et al., 2012; Shaffer et al., 2009) have argued that existing social network models are unable to examine the way in which collective knowledge advances through learner collaboration. Two lines of research have been used to solve this problem. The first approach is to apply a procedure that is similar to an ordinary SNA to examine a different type of social network: a socio-semantic network based on the words that learners use in their discourse. Connections between words are assumed to represent clusters of ideas; the change in the socio-semantic network structure over time is examined both visually and computationally to investigate how a group of students engages in collective knowledge advancement (e.g., Ma et al., 2016; Oshima, Oshima & Fujita, 2018; Oshima et al., 2017). The second approach is another SSNA, based on epistemic frame theory, called Epistemic Network Analysis (ENA) (Shaffer, 2017). ENA is an algorithm that identifies and calculates connections among elements in coded data and visualizes them in dynamic network models that illustrate their structure and strength over time. With ENA, researchers can qualitatively and quantitatively examine cultural practices that participants engage in, such as engineering projects (e.g., Svarovsky, 2011), through their discourse. The present study uses both approaches to investigate different layers of student activity. We have carried out a vocabulary network analysis of their collective knowledge advancement and an ENA to assess their shared epistemic agency.

This study combines the approaches and findings of our previous studies, which examined students' collective knowledge advancement (Oshima et al., 2017), with a further ENA by coding the same discourse data from the perspective of shared epistemic agency.

### Conceptual understanding of complex scientific concepts

Through collective knowledge advancement, learners are expected to engage in complex tasks and comprehension of phenomena. Complex systems are multiple levels of organizations locally interacting with one another, such as financial economies and weather systems (Wilensky & Jacobson, 2013). Studies have revealed that students have difficulty mastering such complex subjects, despite their importance. One reason for the difficulty is that the complex concepts conflict with learners' prior experience. They usually have a "centralized" mindset and tend to favor explanations that assume central control and simple causality. In his interview study, Jacobson (2001) found that undergraduate students were more likely than experts to generate simple causality, central control, and predictability; by contrast, the experts exhibited decentralized thinking about multiple causes, such as stochastic and equilibration processes.

Hmelo-Silver and Pfeffer (2004) have proposed a structure–behavior–function (SBF) framework for assessing different levels of student understanding of complex systems. To assess students' understanding of an aquarium as a complex system, for instance, they used the SBF framework in the following way. Structures are elements of a system; in an aquarium, there are fish, plants, and a filter. Behaviors represent the way in which system structures achieve their purpose; for example, filters remove waste by trapping large particles, absorbing chemicals, and converting ammonia into harmless chemicals. Finally, functions express why an element exists within a given system—in other words, they express the purpose of that system element. For example, the filter removes aquarium byproducts. When researchers studied the verbal responses and pictorial representations of middle-school students, preservice teachers, and experts, they found that novices focused on perceptually available, static-system components. Experts, on the other hand, focused more on interrelations among structures, functions, and behaviors. These results suggest that the SBF framework could be a useful formalism for understanding complex systems.

The present study has used this framework to assess high-school students' conceptual understanding of the human immune system as a complex system. Both before and after the lesson, we asked the students to explain how vaccination protects us from infection by using their knowledge of the human immune system.

## Methods

## Student sample

Thirty-nine tenth-grade students at a high school in Japan participated in this study, as part of their regular curriculum. A science teacher with a Ph.D. in biology and more than ten years of teaching experience taught the students.

## Lesson design

### Structure of the collaborative learning activity: Jigsaw instruction

Twelve groups were formed with three or four students in each and were given a challenge, such as “Can you explain how vaccinations protect us from infections?” They were then provided with three study documents, each of which was needed to solve the challenge. First, individual students from each group gathered to form twelve expert groups (four groups to study each expert material) and worked on their allocated materials over 1.5 lesson periods (each lasting 50 min.). Second, students returned to their original group (the *jigsaw* group) and shared and integrated their knowledge in order to solve the challenge problem. This jigsaw activity took another 1.5 lesson periods. The teacher was responsible for group composition in both group activities.

### Study documents

We created the SBF framework for the human immune system based on a textbook description (Figure 1). We then discussed with the collaborating teacher the way to separate content, during expert group activities, into pieces of knowledge based on three essential functions: *humoral immunity*, *primary and secondary response*, and *cell-mediated immunity*. In the SBF framework, the three functions interact with one another as subsystems.

## Study design

### Pre- and post-tests

Students were questioned individually about their understanding of how the human immune system responds to vaccination (i.e., *Can you explain how vaccination protects us from infection?*). They were given a worksheet with a printed question, on which they could write or draw their ideas. The pre-test was conducted right before the lesson started. The post-test was conducted right after the lesson finished. Each test took one lesson period. Thirty-five students completed both tests; their responses were then analyzed.

### Process data collection

Student conversations during group activities were video-recorded and transcribed for vocabulary network analysis and ENA. We used transcripts of the jigsaw group activity to examine how students exerted their shared epistemic agency in advancing their collective knowledge.

### Analytical procedure

With reference to the SBF framework for the human immune system, student explanatory discourse during the pre- and post-test was categorized into three types: (1) no understanding, (2) single-document understanding, or (3) integrated understanding. When students did not manifest any conceptual understanding of the three documents used in the expert group activity, they were categorized as having no understanding. Students who appropriately demonstrated understanding of one of the three documents were categorized as having single-document understanding. Those who manifested a more complete understanding of the human immune system, based on more than two documents in the expert group activity, were categorized as having integrated understanding. The first and third authors independently evaluated ten randomly selected examples of student discourse in each of the pre- and post-tests. Cohen’s Kappa coefficient for the agreement between the two raters was 0.92. Disagreements were resolved through discussion. The first author evaluated the remaining data. Because all students were categorized as having no understanding in the pre-test, we focused on levels of understanding in the post-test to assess learning outcomes.

To visualize and computationally investigate collective knowledge advancement, Oshima et al. (2017) selected 23 nouns to represent the structures and functions of the human immune system in the SBF framework. On average, students engaged in discourse exchange 358.5 times in the jigsaw groups (SD = 211.8). The socio-semantic network of vocabulary refers to meaningful links (i.e., co-occurrence) between words in exchanges. We then used an application called KBDeX (<http://www.kbdex.net>) to calculate the transition of the total value of degree centralities of nodes in the network across discourse exchanges, following the method used in previous research (Oshima et al., 2012). We then calculated the *term frequencies* (TF) of the selected vocabulary words, using the formula,  $tf(t, d) = 1 + \log(ft, d)$ . We compared TF means across the groups to test whether the amount of

discourse related to the human immune system differed significantly (particularly between high and low learning-outcome groups).

We conducted an ENA of the twelve groups' discourse, using the seven epistemic actions of shared epistemic agency (Damsa et al., 2010) to code each discourse exchange. Cohen's Kappa coefficients for the agreement between the two raters ranged between 0.47 and 1.00. Disagreements were resolved through discussion. Since we were interested in detecting the epistemic frames of shared epistemic agency exerted by all of the groups in relation to their learning outcomes, we conducted separate ENAs for high and low learning-outcome groups.

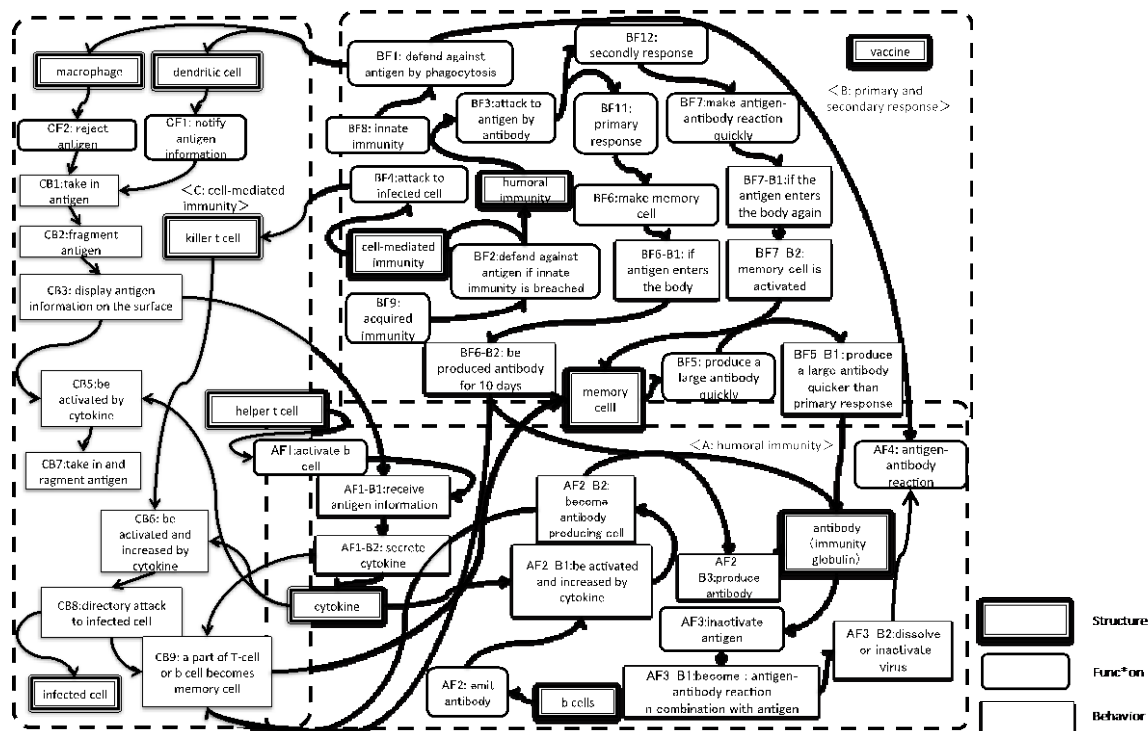


Figure 1. Structure-behavior-function framework of the human immune system.

## Results

### Students' learning outcomes at the individual and group level (Oshima et al., 2017)

We found that twenty-one students integrated SBF understanding. Eleven students demonstrated an understanding of a single part of a learned document; three did not sufficiently learn any piece of the SBF framework. A chi-square analysis of student frequencies across three types of learning outcome showed significance ( $\chi^2 = 13.944$ ,  $df = 2$ ,  $p < .05$ ); in addition, the proportion of students who integrated SBF understanding was higher than the proportion with no understanding. These results suggest that the jigsaw activity did facilitate student integration of knowledge through collaboration, although group differences remained in the learning outcomes. Based on the SBF evaluation, we identified three high learning-outcome groups in which all members acquired integrated conceptual understanding. Nine low learning-outcome groups demonstrated mixed levels of understanding.

### Collective knowledge advancement: the difference between high and low learning-outcome groups

Figure 2 shows the transitions in total values of degree centralities, an index presented in our previous study (Oshima et al., 2017) to detect collective knowledge advancement. We found that the values quickly increased and ultimately exceeded 10.0 in the high learning-outcome groups, while values stayed low and increased slowly across discourse exchanges in the low learning-outcome groups. These results suggest that students in the high-outcome groups engage in collective knowledge advancement more quickly and sustainably.

An additional one-way ANOVA of TFs found no significance among the twelve groups,  $F(11, 198) = 2.35$ ,  $p > .05$ . This result reveals that students in all groups engaged in an equal amount of discourse related to the human immune system.

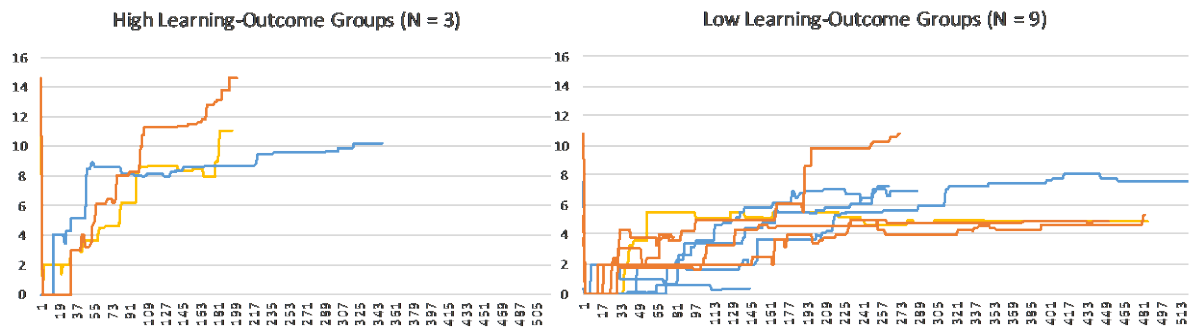


Figure 2. Transitions of total values of degree centralities over discourse exchanges.

### Epistemic frames of shared epistemic agency: the differences between high and low learning-outcome groups

To examine epistemic frames of shared epistemic agency, we conducted an ENA of the discourse of students in high and low learning-outcome groups, using epistemic action codes (Figure 3). Seven codes of shared epistemic agency were plotted in a two-dimensional epistemic space, based on adjacency matrices representing the co-occurrence of the codes. A comparison of code connections in high and low learning-outcome groups revealed the following critical differences. First, we found differences in variances, explained by the two-dimensional epistemic spaces. For high learning-outcome groups, the epistemic space explained 100% of variances in the two-dimensional space (88.9% in the first/horizontal dimension and 11.1% in the second/vertical dimension). By contrast, for low learning-outcome groups, the epistemic space explained only 80% (46.3% in the first dimension and 33.7% in the second). In addition, the relationship between the epistemic and regulative aspects of actions differed in the two types of groups. In high learning-outcome groups, three epistemic actions—alleviating a lack of knowledge (ALoK), creating shared understanding (CSU), and generating collaborative actions (GCA)—were strongly linked to both regulative and projective actions. In low learning-outcome groups, although the three epistemic actions (ALoK, CSU, and GCA) were linked to regulative action, they were only weakly connected to projective action.

In addition to analyzing the differences associated with learning outcomes, we carried out a temporal analysis within each group, dividing the whole process of each group discourse into three phases (Figures 4 and 5). Based on the temporal ENA within each group, we found the following results. First, in high learning-outcome groups, the most active actions changed across the three phases, from ALoK through CSU toward GCA. In each phase, the active action was linked to regulative and projective actions. Low-learning outcome groups, by contrast, began with ALoK and continued to CSU but did not move further toward GCA. ALoK and CSU were linked to regulative action in the first and second phases; they were tied only weakly to projective action.

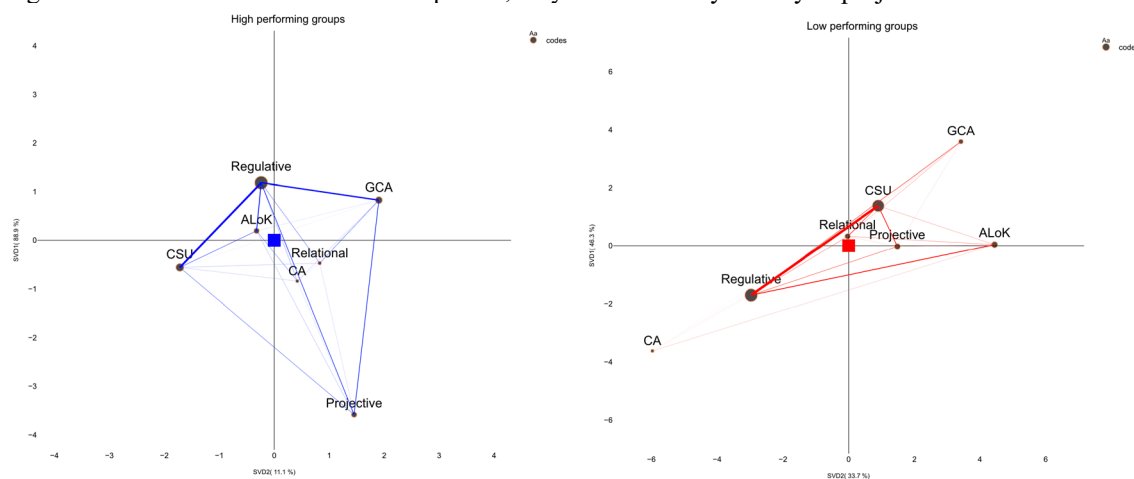
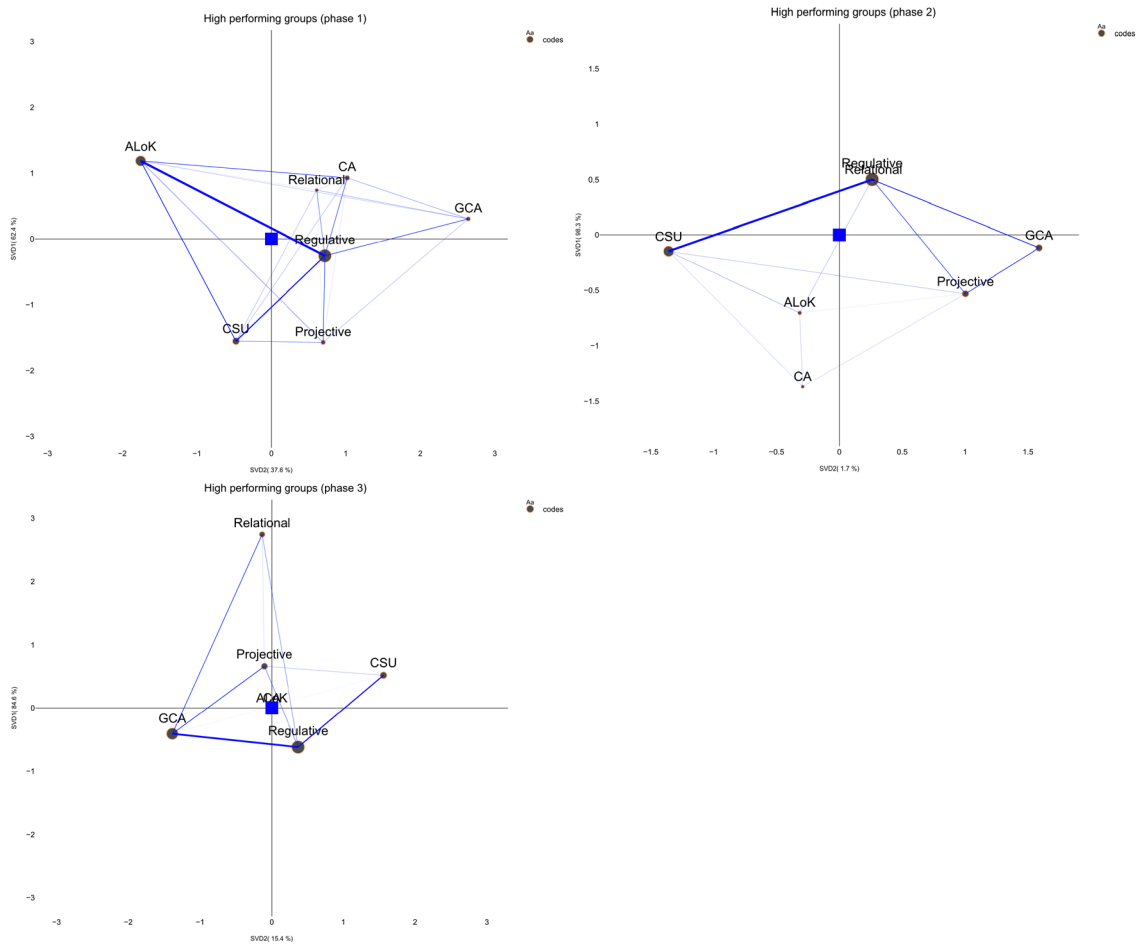
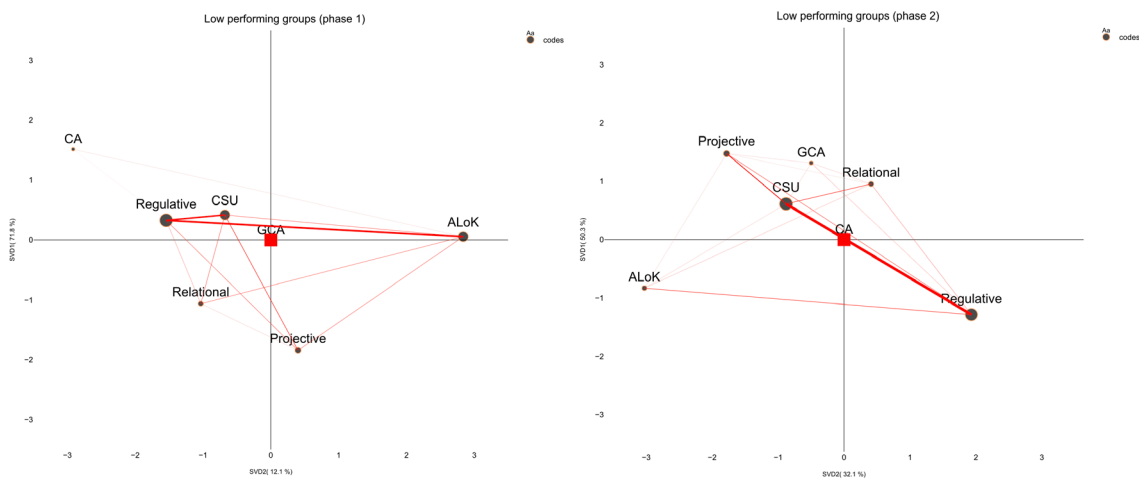


Figure 3. Epistemic frames based on seven epistemic actions in shared epistemic agency—a comparison of high learning-outcome groups (left) and low learning-outcome groups (right).



**Figure 4.** Temporal change in epistemic frames of shared epistemic agency by high learning-outcome groups: first phase (top left), second phase (top right) and third phase (bottom left).



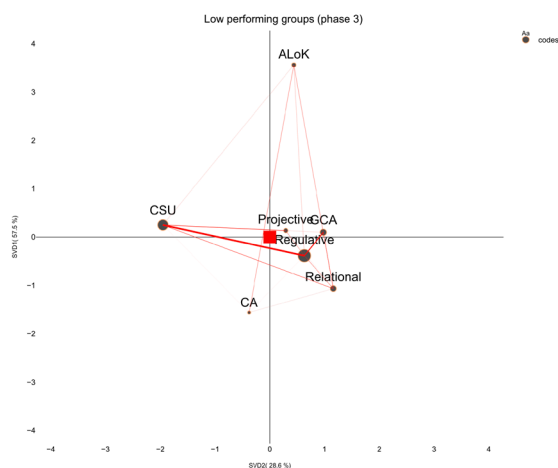


Figure 5. Temporal change in epistemic frames of shared epistemic agency by low learning-outcome groups: first phase (top left), second phase (top right), and third phase (bottom left).

## Discussion

In our previous study, presented at CSCL2017, we looked only at how a jigsaw group activity improved students' ideas involving subject matter knowledge. Although significantly more students succeeded in acquiring integrated SBF understanding, we also found big differences in learning outcomes among student groups. Only three of the twelve groups (25%) succeeded in their collaboration, a problem that has now been examined more systematically using a double-layered socio-semantic network analysis. This approach has allowed us not only to analyze what happened in the jigsaw group activity (collective knowledge advancement) but also how it happened (shared epistemic agency). The new findings, mainly obtained through an ENA, led to the following interesting interpretations of the differences between high and low learning-outcome groups.

First, our additional analysis of vocabulary term frequencies related to the SBF framework for the human immune system revealed no significant difference in the amount of subject-related discourse across the groups. The fact that the amount of discussion was the same suggests that learning outcomes were not related to how much the students talked about the study topic. Instead, the way in which they scrutinized their own ideas appears to have been the key to more productive group-learning outcomes.

Second, our ENA of students' epistemic actions showed that the two types of groups (high and low) engaged in different types of epistemic agency. The high learning-outcome groups established more robust epistemic space (100% of explained variances) than the low learning-outcome groups (80% of explained variances). The most epistemic aspect of actions was linked to the regulative aspect of actions in the high learning-outcome groups. In addition to constructing a shared understanding, they also collaboratively generated actions based on their shared understanding to explain how vaccination protects us from infection. They took regulative and projective actions in relation to epistemic actions. In other words, they monitored their understanding and attempted to improve their ideas continuously, in order to create their own theories of the human immune system. No strong link between epistemic and regulative actions was found in the low learning-outcome groups.

Finally, our temporal ENA of the students' epistemic agency provided a more accurate understanding of the way in which they engaged in shared epistemic agency to advance collective knowledge. The high learning-outcome groups were more engaged in alleviating their lack of knowledge during the first phase of activity—before they constructed a shared understanding. By contrast, the low learning-outcome groups engaged in two types of epistemic actions simultaneously, starting in the first phase. During the third phase, they continued to construct a shared understanding and did not take the next step toward generating collaborative actions (for example, by creating new collaborative ideas). These results suggest that there may be an optimal sequence of epistemic actions, such as alleviating a lack of knowledge before constructing a shared understanding. The high learning-outcome groups appeared to monitor themselves and proceed in accordance with a sequential process.

By analyzing learning outcomes, collective knowledge advancement, and shared epistemic agency, we were able to see how students engaged in collective knowledge advancement through their shared epistemic agency when successfully acquiring a deep conceptual understanding. As previous research (e.g., Damşa, 2014; Scardamalia, 2002) suggests, epistemic agency plays a key role in successful collaboration. The present study introduces a new finding: epistemic actions in collective knowledge advancement are likely to follow a sequential

process. This hypothesis should be further tested in future studies, which could incorporate design elements into jigsaw instruction to help students become aware of the sequence.

## References

- Bereiter, C. (2002). *Education and mind in the knowledge age*. Hillsdale, NJ: Erlbaum.
- Brown, A. L. & Campione, J. C. (1996). Psychological theory and the design of innovative learning environments: On procedures, principles, and systems. In L. Schauble & R. Glaser (Eds.), *Innovations in learning: New environments for education*. (pp. 289–325). Mahwah, NJ: Lawrence Erlbaum Associates.
- Damşa, C. I. (2014). The multi-layered nature of small-group learning: Productive interactions in object-oriented collaboration. *International Journal of Computer-Supported Collaborative Learning*, 9, 247–281.
- Damşa, C. I., Kirschner, P. A., Andriessen, J. E. B., Erkens, G., & Sins, P. H. M. (2010). Shared epistemic agency – An empirical study of an emergent construct. *Journal of the Learning Sciences*, 19(2), 143–186.
- Hmelo-Silver, C. & Pfeffer, M. G. (2004). Comparing expert and novice understanding of a complex system from the perspective of structures, behaviors, and functions. *Cognitive Science*, 28, 127–138.
- Jacobson, M. J. (2001). Problem solving, cognition, and complex systems: Differences between experts and novices. *Complexity*, 6(2), 1–9.
- Ma, L., Matsuzawa, Y., Chen, B., & Scardamalia, M. (2016). Community Knowledge, Collective Responsibility: The Emergence of Rotating Leadership in Three Knowledge Building Communities. In C. K. Looi, J. L. Polman, U. Cress, & P. Reimann (Eds.), *Transforming Learning, Empowering Learners: The International Conference of the Learning Sciences (ICLS) 2016*, Volume 1. Singapore: International Society of the Learning Sciences.
- Miyake, N. (1986). Constructive interaction and the iterative process of understanding. *Cognitive Science*, 10, 151–177.
- Miyake, N. & Kirschner, P. A. (2013). The social and interactive dimensions of collaborative learning. In K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences, the Second Edition*. (pp. 418–438). NY: Cambridge Univ. Press.
- Oshima, J., Ohsaki, A., Yamada, Y., & Oshima, R. (2017). Collective knowledge advancement and conceptual understanding of complex scientific concepts in the jigsaw instruction. In Smith, B. K., Borge, M., Mercier, E., and Lim, K. Y. (Eds.), *Making a Difference: Prioritizing Equity and Access in CSCL, 12th International Conference on Computer Supported Collaborative Learning (CSCL) 2017, Volume 1* (pp. 57–64). Philadelphia, PA: International Society of the Learning Sciences.
- Oshima, J., Oshima, R., & Fujita, W. (2018). A Mixed-Methods Approach to Analyze Shared Epistemic Agency in Jigsaw Instruction at Multiple Scales of Temporality. *Journal of Learning Analytics*, 5(1), 10–24.
- Oshima, J., Oshima, R., & Matsuzawa, Y. (2012). Knowledge Building Discourse Explorer: A social network analysis application for knowledge building discourse. *Educational Technology Research & Development*, 60, 903–921.
- Paavola, S., & Hakkarainen, K. (2005). The knowledge creation metaphor—An emergent epistemological approach to learning. *Science & Education*, 14, 535–557.
- Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Ed.), *Liberal education in a knowledge society* (pp. 67–98). Chicago, IL: Open Court.
- Scardamalia, M., Bransford, J., Kozma, B., & Quellmalz, E. (2012). New assessments and environments for knowledge building. In P. Griffin, B. McGraw, & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 231–300). New York, NY: Springer Science+Business Media.
- Shaffer, D. W. (2017). *Quantitative ethnography*. Madison, WI: Cathcart Press.
- Shaffer, D. W. et al. (2009). Epistemic network analysis: A Prototype for 21st Century assessment of Learning. *International Journal of Learning and Media*, 1(2), 1–21.
- Svarovsky, G. N. (2011). Exploring complex engineering learning over time with epistemic network analysis. *Journal of Pre-College Engineering Education Research*, 1(2), 19–30.
- Wilensky, U. & Jacobson, M. J. (2013). Complex systems and the learning sciences. In K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences, the Second Edition*. (pp. 319–338). NY: Cambridge Univ. Press.

## Acknowledgments

We thank the participating teacher and students. This work was supported by JSPS KAKENHI Grant Number 16H0187.

# Families and Media Multi-tasking: Reorganizing Collaborative Learning at Home

Deborah Silvis, University of Washington, dsilvis@uw.edu

**Abstract:** As the number of tools available for learning within computer-based tasks proliferates, so too do the tasks. This paper describes how families use technology and new media for multi-tasking in order to collaboratively accomplish many routine tasks around their homes. Families employ a number of approaches to media multi-tasking, and their sociotechnical practices reflect how children learn *to* multi-task, learn *through* multi-tasking, and learn *despite* multi-tasking. I draw on these practices to explain how families manage shared- and divergent -objects of activity at home, and how technology is implicated in this maintenance work. I offer a multi-dimensional framework for multi-tasking that suggests how determining what is consequential for learning with technology is often contingent on a number of tools, tasks, purposes, and people.

## Introduction

On a bright April day, after nine-year-old twins Oscar and Eddie had returned home from school, I interviewed their mom Steph about her sons' technology routines. After detailing their "daily media round", or all the routine places people go and things they do with media and technology (Taylor, Takeuchi & Stevens, 2017), Steph described some of the activities they did as a family on a more infrequent basis. For example, she participated in several community organizations, and the twins usually accompanied her when she attended weekly or biweekly meetings. Steph gave the following account of her sons' involvement during such family outings.

If they want to take off the headphones and engage in the dialog at the meeting, they can participate and soak things in. Sometimes in the car on the way home they start to talk about what the adults were discussing. Even though they are [playing] on their tablets or with their headphones, they are still soaking in what's going on around them. *They are obviously multi-tasking*, because they will reengage in the conversation [emphasis added].

Later in this paper I will describe how the twins (and their mom) accomplished such multi-tasking, a phenomenon I observed across all the families I studied. This paper examines how families' approaches to media engagement allow learners to "soak in what's going on around them" despite multiple demands on their attention.

In any given day, families must accomplish meal prep and eating, clean-up and chores, commuting, attending school, paid work, homework, community meetings and parties, shopping, sports or music practice, paying bills, worship, play time, pet care, bedtime, and more. The work is never done, and in today's socioeconomic context, families are busier than ever (Ochs & Kremer-Sadlik, 2013). But, as Graesch (2009) suggested, "busier families are not necessarily working harder or longer hours, but instead are having to reconfigure their lives around a growing number of bids for their attention" (p. 86). How is it that families manage to do more in the same amount of time? How is technology involved in reconfiguring family life? What material and human resources must they collaboratively assemble to manage it all? This paper examines multi-tasking as one response to these questions and explores how it is part and parcel of learning in new media ecologies (Barron et al., 2009).

The current paper represents a condensed version of a longer dissertation chapter analyzing media multi-tasking in the context of family life. In what follows, I address the question: *what forms of collaborative activity does multi-tasking make possible, and how do families collaboratively contribute to what ends being consequential for learning?* Drawing on three instances of families' media multi-tasking, I interrogate school-like notions of being "on task" and instead take the perspective that multi-tasking is intrinsic to most everyday activities. I elaborate dimensions of multi-tasking that bear on learning: how tasks involve arranging for multiple sub-tasks, multiple tools, multiple purposes, and multiple people. Analyzing instances of families' multi-tasking across these dimensions showed how families collaboratively organize activities to learn *through*, learn *to*, and learn *despite* multi-tasking.

At stake across all these elements of multi-tasking is a question of whose perspective on a task matters



for learning. Children and adults have different views of what they want to accomplish during precious hours spent at home (Goodwin & Cekaite, 2018). Likewise, their objectives within a shared task can vary; adults may see media engagement as useful for one purpose, while for children, it serves an entirely different aim. Children frequently choose the course of their media engagement; however, doing so in ways that conform to families' shared rules and norms can necessitate children's multi-tasking. Multi-tasking then becomes a lens through which what is consequential *for learners* becomes visible, opening up endogenous perspectives with the potential to expand what counts as meaningful media engagement and learning and for whom (Hall & Jurrow, 2015; Stevens, 2000). Viewed in this way, routine tasks are sites of on-going negotiation in families, where multi-tasking can resolve- or renew- tensions over technology use. As the current analysis will show, home is a powered place, where decisions about sociotechnical practices are charged and pivot around multiple tools, people, objectives, and tasks.

## Conceptual framework

According to some researchers, young people are both the biggest multi-taskers and the ones least likely to perform well while multi-tasking (Rosen, Carrier & Cheever, 2013). Because multi-tasking performance has consistently been found to improve with age, teaching and learning has been a major focus of multi-tasking research. And perhaps unsurprisingly, in recent years, given the proliferation of mobile media and screen-based learning designs, there has been renewed interest in multi-tasking research. Lin and colleagues (2011) have produced a number of studies of media multi-tasking (MMT), which they define in terms of dual-tasking, as "engaging in multiple media and tasks simultaneously" (P. 183), for example reading a news story while watching a video (Lin, Mills, Ifenthaler &, 2016). The bulk of literature has pointed to negative effects of media multi-tasking on studying at home (Rosen et al., 2013), classroom learning (Wood et al., 2012), and social success of young people (Pea et al., 2012).

Studies of families' multi-tasking leave the distinct impression that children either *are* primary distractions (for parents) or they are primarily distracted *by* media (Craig, 2016). While there may be some truth to this gloss on family multi-tasking, it neglects how children's learning might be involved in new forms of media multi-tasking or how learning may be supported through collaborations with other family members (Shapiro et al., 2017; Silvis, Taylor & Stevens, 2018b; Lin et al., 2016). Recently, Mejía-Arauz and colleagues (2018) have examined collaborative learning at home, distinguishing between task objectives that are met through "negotiation," completed individually and then combined together, and those achieved "collaboratively," through joint synchronous activity. This work points to how families' collaborative activities frequently exhibit multi-tasking during culturally valued routines like cleaning, cooking, care-giving, or conversing, though this was more normative in families from non-dominant backgrounds, such as Mexican heritage families.

In other research with families, joint media engagement (JME) may also provide a supportive collaborative context for multi-party media multi-tasking (Takeuchi & Stevens, 2011). Studies of JME have by and large looked at how multiple people arrange for learning around *single* devices or media, which are generally digital. However, in today's dynamic media ecologies, digital technologies are not the sole means of multi-tasking (e.g. Andrade, 2010). A significant amount of paper-based activities persist in homes and are interleaved with screen-based activities. Therefore, multi-tasking not only involves doing multiple tasks at once, it also represents a *multi-tooled* approach to tasks. As collaborators, whose labor is distributed, draw these materials together in-task, the task itself may break down into multiple sub-tasks, with separate objectives that must be articulated and coordinated (Stevens, 2000).

The nature of task objectives is a central issue in research on multi-tasking. A single task may achieve multiple aims (Engeström, 2001). Take for example a routine activity like checking email. Because many different topics and tasks sit within different email messages, the objects of an activity such as "checking email" can range from coordinating a meeting with colleagues to getting the latest deals on produce from the local grocery store. Task objectives proliferate in complex media environments, even within a single task. Engeström suggests how "the object of activity is a moving target, not reducible to conscious short-term goals" (p. 136). There is no "task" only "task(s)" as people and objects co-construct them. Or, as Erickson (2016) has commented, "in any human interaction, there is always more than one thing going on at the same time." Relatedly, activities that serve one person's objectives may engage someone else in the family for a variety of different reasons, only some of which are shared among collaborators. Managing interactional "floors" (i.e. interactional spaces for establishing topics or tasks), then requires balancing dynamic tasks and topics across time and space (Erickson, 2004). Multi-tasking,

therefore, is not only a means of accomplishing multiple discreet *tasks* or sub-tasks with multiple *tools*, it is also *multi-party* and responds to the *multiple purposes* of different participants.

## Study design and methods

In this analysis, I take a situated view of learning (Brown, Collins & Duguid, 1989; Lave & Wenger, 1991) and draw on data collected in an ethnographic study of family life at home as it is currently being transformed by new (especially mobile) forms of media and technology. Participants included eighteen focal children in twelve families from diverse racial, ethnic, geographic, and socioeconomic backgrounds. Participants were recruited from local youth-serving organizations, camps, and other places where young people between the ages of nine and thirteen years old spend time; this period of development is significant for media engagement because it is about the time when children get their own devices (Rideout & Katz, 2016).

Data collection in this ethnographic study of families “daily media rounds” (Taylor, Takeuchi & Stevens, 2017) took place over two years in two separate US cities, and methods included the following: semi-structured interviews with parents and children (48 total); video recorded observations conducted during home visits, some of which were recorded by children using point-of-view cameras (i.e. GoPros®) (approx. 100 hrs); experience sampling through nightly phone calls (90 total calls), and a novel research activity for digital mapping of participants’ technology use (Silvis, Taylor & Stevens, 2018a), which was also video recorded (16 digital artifacts). To answer the research questions- how families’ media multi-tasking organizes learning arrangements- I drew on ethnographic and interaction analysis (Jordan & Henderson, 1995). I produced multi-modal transcription of the talk, gesture, gaze, coordination of body movements (or lack thereof), uptake of tools, use of space, and other aspects that played into moment-to-moment interactions (Tulbert & Goodwin, 2011).

## Analytic findings

A wide variety of everyday tasks took place across families participating in the study. I analyzed a number of key instances where multiple people, tools, tasks, and aims were being configured and re-configured, as opposed to those where a child was independently engaged with a single task for a period of time (i.e. reading a book or e-book for thirty minutes without interruption, doing math homework at the kitchen table before moving on to play time). Looking across tasks that were multi-tooled, multi-party, multi-purpose and involved multiple sub-tasks, a number of patterns emerged. I found that families approaches to multi-tasking supported children learning *through* multi-tasking, learning *to* multi-task, and learning *despite* multi-tasking. These patterns reflect the dynamics of multiple dimensions of multi-tasking in-task (i.e. multiple tools, people, tasks, and aims). In what follows, I focus on three instances of children multi-tasking, emphasizing the role of collaboration in each case. Each of these examples serves as a representative case of three ways in which multi-tasking relates to learning.

### Learning *through* multi-tasking using a single device

Because wired digital technologies afford endless possibilities for media engagement, a single device can organize a space for a great deal of media multi-tasking. Katherine, a ten-year-old girl whose family had recently moved to a large city in the Northwest US from Shanghai, was rarely without her prized possession, a MacBook Air. According to her mom, when she was not participating in a summer program, taking ballet lessons, or making plans to attend her new school in the fall, Katherine was inseparable from this mobile device. She relied on the tool for a myriad of activities including: *tasks assigned by her Mom* (e.g. researching animal shelters where they could adopt a family dog or places where they could go to explore their new city); *homework assigned by teachers* (e.g. mythology reading, poetry writing); *preparing for school in the fall* (e.g. checking summer reading list, studying for a gifted program placement test); *learning about topics of interest* (e.g. receiving a video from a friend in her Kitchen Chemistry class, receiving email from a list serve about “science-y” cooking); and *independently surfing websites for news or videos* (e.g. watching an old episode of Pokémon on YouTube, reading articles in her news feed).

During one visit to her home, I observed how Katherine used her laptop over the course of an hour for a number of separate and simultaneous activities, a practice I call “single-device multi-tasking.” Zeroing in on this observation alone illustrates the density of media multi-tasking Katherine engaged in during more than an hour spent using her laptop. As she worked, a single device provided the means of easily multi-tasking by tabbing back and forth between active windows, continuously checking the status of uploads, or opening new windows and programs in order to accomplish a task at hand. When I joined her in her bedroom work site on this day, Katherine was already busy working on a video she had animated in QuickTime the previous evening. She was uploading it to a site called Magisto.com, that hosts videos for sharing or publishing on social media. While this underlying activity progressed in the background, she busied herself with a number of other tasks (Figure 1).

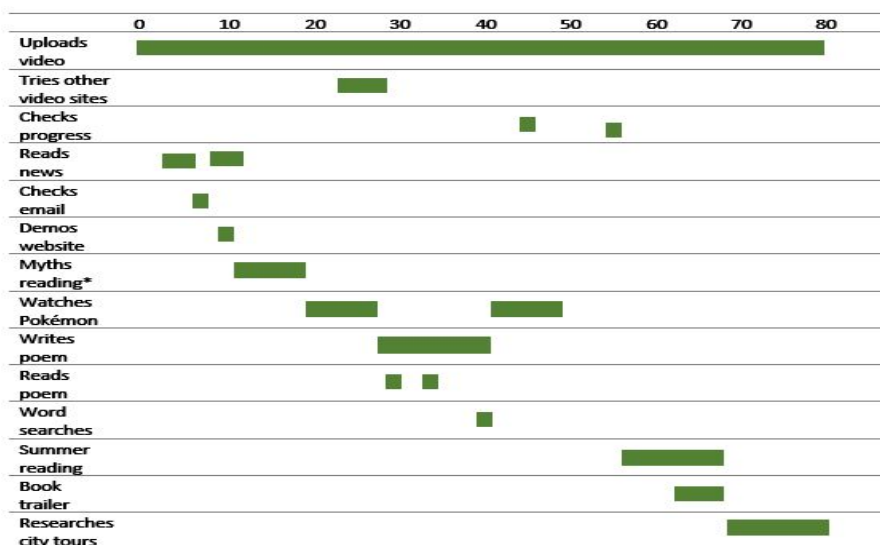


Figure 1. Single device multi-tasking. Timeline (in minutes) of all the separate, overlapping, tasks Katherine engaged in during a single observation. \*Except for Mythology reading, all of these tasks utilized only her laptop.

After about thirty minutes (and at least seven discreet tasks), Katherine began composing a poem in a Google Doc that she shared with her summer program writing teacher, who provided her with virtual feedback. A few minutes into her poetry writing, Katherine opened Google Images in a separate tab, something she said she sometimes did “for inspiration.” She referred to several images of poems, tabbing back and forth between these and the one she was authoring, and eventually positioning multiple windows in order to simultaneously view multiple poems on screen. When she hit an impasse in her writing, she opened up another tab to search on Thesaurus.com, which she identified as her “favorite website.” In the course of writing the poem, she managed the display by juxtaposing- and tabbing back and forth between- these multiple tabs and windows. Configuring multiple resources on-screen enabled her to navigate sub-tasks in a way that supported learning through multi-tasking.

After working on the poem for fifteen minutes, she resumed watching an episode of Pokémon on YouTube that she had interrupted to write the poem, and later moved on to do some research on *Animal Farm*, which was on her summer reading list for school next fall. However, she kept open the multiple tabs she had used to write the poem, so it was possible to return to the task later on. Katherine’s multi-tasking led her across multiple sites and platforms to engage in multiple tasks using a single device. She was able to learn about poems, literature, vocabulary, and much more through her multi-tasking and also with the help of virtual collaborators. In the course of my observations, she interacted with her friends, her mom, and her dance teacher via email. At any given time, a wired laptop gives young people multiple avenues for interaction, and who they choose to connect to shifts as the focal task changes. Consequently, learning through single-device multi-tasking is not adequately characterized as a solo activity, nor does it simply serve as a means of individual efficiency or personal productivity. Media multi-tasking is mutually monitored by multiple collaborators (Goodwin & Cekaite, 2018)- albeit virtually and asynchronously. Even when a single person uses a single device, multi-tasking demands- and is a response to demands for- collaborative learning across tools, people, and tasks.

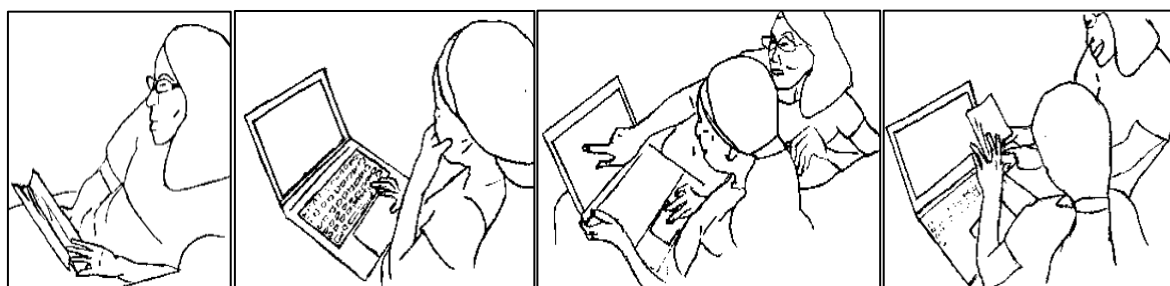
### Learning to multi-task by mixing tools and negotiating objectives

For Katherine, a single tool became a site for organizing learning through multi-tasking across heterogeneous objects of activity. Conversely, it was often the case for the families I studied that a single *task* became a site for learning to multi-task across multiple tools and task objectives. In today’s media ecologies, families have many options for equipping and jointly accomplishing routine tasks. Often, family members’ purposes for performing tasks vary and need to be negotiated amongst people and devices. Placing a multi-tasking lens on a single task can surface the often hidden sociotechnical work required to stabilize a task so that people can jointly accomplish their aims with multiple technologies (Star, 1990; Stevens, 2000). Learning to multi-task becomes necessary when task-objectives diverge, when tools are distributed, or when one person delegates their task to another.

Such was the case when Natalie, an eleven-year-old who lived with her grandmother and mother Gina near a

large Midwestern US city, helped her mom complete an Excel spreadsheet containing information from a survey Gina had distributed at work at a local service organization. Gina introduced the task to Natalie as work they could do together, enlisting her daughter in her own work as a valued coparticipant. Barron and colleagues (2009) referred to this arrangement as “parents as employers,” a situation that arises when parents “entrust their technologically skilled children to perform technical services for them” (p. 69). Rifling through the questionnaires, Gina told Natalie that they needed to make a spreadsheet and that Natalie could do the data entry. When Natalie was unsure of what this meant and feigned resistance, claiming it was a weekend and she shouldn’t have to work, her Mom reassured her by playfully smacking her arm and announcing that Natalie was going to learn something new. What for adults may seem mundane (e.g. building a spreadsheet, doing data entry) presents productive challenges and novel tasks for young learners, who may have very different task objectives and motives than their adult collaborators.

It is also important to note here that Natalie was recruited *by her Mom* to do this work, which is a different starting point for the delegation of young people’s computer-based work that typically originates outside the home in school. As opposed to homework assigned by a teacher, where a single task is undertaken for the putative purpose of completing an assignment (or earning a grade), building an Excel spreadsheet was more multi-purpose; this contributed to the pair approaching their task with divergent objectives. Gina reported that she often recruited Natalie into tasks for her job if technology was involved, in order to teach her daughter new technical skills. For Natalie, spending time with her Mom during their free time on a Saturday was a powerful motive for completing data entry, a task that adults may view as thankless grunt work and some children might regard as a pointless chore. Learning to build a spreadsheet entailed learning to negotiate multiple people’s objectives in-task and to coordinate their separate sub-tasks. Therefore, in addition to learning to use Excel, Natalie was also learning to multi-task.



**Figure 2.** Multiple tools within a task. (Left to right) Natalie’s mom reports information from paper-based surveys; Natalie enters data into Excel spreadsheet; they collaborate over the screen; they collaborate over paper.

Learning *to* multi-task meant drawing on multiple material means of mediation as well. As they began the spreadsheet together on the couch in the living room, Gina held the paper surveys, while Natalie was responsible for computer data entry (Figure 2). Despite the proliferation of digital means of doing domestic tasks, there is continued need to mix multiple types of tools together to accomplish a task (Silvis et al., 2018b), and paper is still a widely used resource for collaborative knowledge work (Brown & Duguid, 2017). The mixing of technologies within a single task was pervasive in children’s activities across all the families we observed; this spreadsheet multi-tasking literally spread the task out across sheets of paper, software, and screens and fractured the single task into sub-components made visible through an emergent division of labor. Learning to build a spreadsheet meant learning to multi-task by drawing on- and drawing together- multiple material means and multiple learning objectives.

### Learning *despite* multi-tasking as domestic complexity multiplies

The complexities of maintaining everyday routines and organizing family life are compounded when multiple people employ multiple technologies across multiple tasks, timeframes, and spaces. A family of three, with jobs, hobbies, school, and close family who lived nearby, multi-tasking was typical in the Hernandez household. This was exemplified one day after school when nine-year-old twins Eddie and Oscar completed a homework assignment with their mother’s assistance. In this example, multi-tasking was multiplied through the complex spatiotemporal and material arrangements, whereby people accomplish a number of different tasks together at home. When I arrived to observe them one day after school, Oscar and Eddie were already immersed in their typical after-school routine: watching TV, completing homework, and playing video games. Their homework

assignment on this particular day was an on-going project in their language arts class that asked them to create a book cover for their favorite book. Both Eddie and Oscar had chosen to create book covers based on *Pokémon Deluxe Essential Handbook*, an encyclopedia of over seven hundred Pokémon, thoroughly dog-eared and annotated by the twins, who each had their own copy.

Rounds of working on homework, which the boys completed primarily at the kitchen table, were punctuated by “technology breaks,” a strategy their family implemented in the afternoons each day after school. The twins’ breaks occurred at different times, creating a dynamic task environment in which tasks were not entirely distinct. For example, the TV stayed on in the background while they worked on their homework, compelling the twins to engage in dual-tasking and splitting their attention during tasks (Lin et al., 2011). For his part, Eddie designed and colored his book cover illustration while watching cartoons, and later continued watching TV while taking a technology break and playing *Angry Birds* on his tablet. However, with multiple people engaged in multiple tasks simultaneously, multi-tasking quickly multiplied. The living room was usually reserved for technology breaks, but Steph made an exception so that Oscar could research a Pokémon illustrator for the back flap of *his* book cover at a moment when Eddie’s work crowded the kitchen table. Oscar and Steph sat comfortably together in a side-by-side configuration, co-viewing her laptop screen while he worked (and stole glances at the TV), a prototypical configuration of bodies during JME (Takeuchi & Stevens, 2011).

As in any complex learning environment, there are always multiple things going on at once (Erickson, 2016). While the twins did their homework, their mom was busy coordinating the delivery of new furniture for their bedroom. Steph had told the boys that she wanted them to each tip the movers one dollar for assembling their furniture, and she placed two one-dollar bills in the threshold between the kitchen and living room, directly in the pathway the twins would likely travel (Figure 3). She likened this to tipping the pizza guy, and she emphasized to them that “I want you to tip them because they’re providing a service.” The boys meanwhile remained unresponsive and continued gazing at the TV from the kitchen table where they worked. A short time later, Oscar and Eddie raced down the hall after snatching up the bills, and they diligently tipped the movers who subsequently departed. Mom immediately noticed that the movers had left their drill in the entrance hallway and chased after them down the driveway. Eddie wryly commented “maybe that is a tip for *us*,” referring to the drill. This re-voicing of his mother’s task directive indicates how he had heard and understood her words in a way that allowed him to recycle them in an entirely different context (Goodwin, 2018). Whereas, several weeks prior, Steph had reported how the twins were able to “soak things in” despite their multi-tasking, in this instance they demonstrated how this looks in practice. Despite their media multi-tasking, Eddie and Oscar learned valuable lessons about what it means to contribute to household routines and collaborations, where multiple tasks vie for people’s attentional resources.



Figure 3. Multiplying media multi-tasking. Mom and Eddie both discuss “tipping” the delivery guys, despite the demands of multiple tasks, people, and tools in their home media ecology.

## Conclusions and implications

Across these three cases, I have shown how media multi-tasking supports learning at home and some forms that multi-tasking can take, including *single-device multi-tasking*, *mixing materials and multiple purposes within a single task*, and *multiplying multi-tasking* by adding more tasks and people to the task setting. I have used these examples to elaborate how children learn *through*, learn *to*, and learn *despite* multi-tasking. Families’ designs for learning at home support these forms of multi-tasking, and this takes sustained interactional work and

intergenerational collaboration. Media multi-tasking is common in families, and we can see vividly how multi-tasking multiplies quickly as people move themselves and their tasks around their homes. Even in the case of Katherine and Natalie, relatively stable (seated) sites of single tasks can quickly expand into sub-tasks using networked devices.

Prevailing perspectives on multi-tasking caution against it, for how it distracts learners, divides attention (and families), and threatens successful task completion. As a response, this analysis contributed to a different perspective on multi-tasking: media multi-tasking is a significant and pervasive sociotechnical practice, and children engage in many forms of multi-tasking in diverse learning environments. I have suggested a multi-dimensional model for multi-tasking that supports analysis of the multiple tasks, tools, objectives, and people entailed by many everyday learning activities. Understanding how these are multiply configured from multiple perspectives contributes to an expansive view of multi-tasking, widening the lens on what counts as valuable learning arrangements, where and when this takes place, and who decides.

Whereas conventional multi-tasking research finds that adults are better than children at multi-tasking and then uses this to justify cautionary narratives about children's media multi-tasking, I am arguing that adults are important supporters of learning through, to, and despite children's multi-tasking. Because most of us *do* frequently multi-task, it is worth considering how children come to be multi-taskers and to design for this. Simple exposure to multiply mediated learning environments is insufficient for learning to multi-task; Natalie was quite proficient using her laptop but needed her Mom's assistance to learn to multi-task on it. As with children's development of other technological competencies, learning to multi-task benefits from parental support, and is a sociotechnical achievement (Barron et al., 2009).

In the current analysis, I treated home environments as especially conducive to learning while multi-tasking, and I focused on families' designs for multi-tasking. In response to anxious or nostalgic narratives that eschew multi-tasking as a valued learning practice, I believe much can be learned from understanding families' approaches to learning through, to, and despite multi-tasking. At the same time, I do not see these three approaches to multi-tasking as entirely distinct. It was often the case that learning to manage multiple tabs on a screen, for example, also involved learning about whatever content was contained in those windows. Furthermore, families' multi-tasking organized learning such that background processes- like watching TV, monitoring other family member's activities, or uploading a video- intermittently became the central focus of activity. This suggests that task objectives at home are better characterized as moving targets than as categorically "focal" or "background" tasks.

Distinguishing background from foreground is a paradigmatic problem of perspective. What one person (i.e. a parent or teacher) thinks is important to do at any given time may not be valued by children and vice versa. Keying into how tasks serve multiple purposes points to how assigning and distributing tasks over time, space, people, and materials is frequently a powered practice. While school-like versions of attention often emphasize individuals "staying on task," home media ecologies are rich sites for multi-party multi-tasking where the objectives of tasks- and their means of completion- originate from more equitably distributed starting points. My analysis supports the suggestion by Lin and colleagues (2016) that collaborative learning is enhanced by multi-tasking because of the excitement of engaging in more than one task at once, a learning configuration often forbidden in formal learning settings. Instead of barring learners from using multiple media and devices to accomplish tasks in school, designs for learning might incorporate more people, technologies, and tasks in order to simulate how children immerse themselves in dynamic home media ecologies and to stimulate interest-driven learning.

## References

- Andrade, J. (2010). What does doodling do? *Applied Cognitive Psychology*, 24:100-106.
- Barron, B., Martin, C. K., Takeuchi, L., & Fithian, R. (2009). Parents as learning partners in the development of technological fluency. *International Journal of Learning and Media*, 1(2), 55-77.
- Brown, J.S., Collins, A., Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, 18(1), 32-42.
- Brown, J.S. & Duguid, P. (2017). *The social life of information*. Boston, MA: Harvard Business Review Press.
- Craig, L. & Jenkins, B. (2016). The composition of parents' and grandparents' child-care time: gender and generational patterns in activity, multi-tasking, and co-presence. *Aging & Society*, 36, 785-810.
- Engstrom, Y. (2001). Expansive learning at work: Toward an activity theoretical reconceptualization. *Journal of*

- education and work*, 14:1, 133-156.
- Erickson, F. (2004). *Talk and social theory: Ecologies of speaking and listening in everyday life*. Polity Press.
- Erickson, F. (2016). Learning to Look and Listen: Group Viewing Session. *Spencer Foundation Conference*, Arizona State University.
- Goodwin, C. (2018). *Co-operative action*. NY: Cambridge University Press.
- Goodwin, M.H. & Cekaite, A. (2018). *Embodied family choreography: Practices of control, care, and mundane creativity*. London: Routledge.
- Graesch, P.A. (2009). Material indicators of family busyness. *Social Indicators Research*, 93, 85-94.
- Hall, R. & Jurow, S. (2015). Changing concepts in activity: Descriptive and design studies of consequential learning in conceptual practices. *Educational Psychologist*, 50:3, 173-189.
- Jordan, B. and Henderson, A. (1995). Interaction analysis: Foundations and practice. *The Journal of the Learning Sciences* 4 (1), 39-103.
- Lave, J. & Wenger, E. (1991) *Situated learning: Legitimate peripheral participation*. Cambridge, UK: Cambridge University Press.
- Lin, L., Lee, J. & Robertson, T. (2011). Reading while watching video: The effect of video content on reading comprehension and media multitasking ability. *Journal of Educational Computing Research*, 45(2).
- Lin, L., Mills, L.A. & Ifenthaler, D. (2016). Collaboration, multi-tasking and problem solving performance in shared virtual spaces. *Journal of Computers in Higher Education*, 28, 344-357.
- Mejía-Arauz, R., Rogoff, B., Dayton, A., and Henne-Ochoa, R. (2018). Collaboration or negotiation: Two ways of interacting suggest how shared thinking develops. *Current Opinion in Psychology*, 23:117-123.
- Ochs, E. & Kremer-Sadlik. (2013). Introduction. In E. Ochs & T. Kremer-Sadlik (Eds.), *Fast-forward Family: Home, Work, and Relationships in Middle-Class America*. Berkeley, CA: University of California Press.
- Pea, R., Nass, C., Meheula, L, Rance, M, Kumar, A., Bamford, H... & Zhou, M. (2012). Media use, face-to-face communication, media multitasking, and social well-being among 8-12-year-old girls. *Developmental Psychology*, 48:2, 327-336.
- Rideout, V. & Katz, V. (2016). *Opportunity for all? Technology and learning in lower-income families*. A report of the Families and Media Project. New York: The Joan Ganz Cooney Center at Sesame Workshop.
- Rosen, L.D., Carrier, L.M. & Cheever, N.A. (2013). Facebook and texting made me do it: Media-induced task-switching while studying. *Computers in Human Behavior*, 29, 948-958.
- Shapiro, B., Hall, R., & Owens, D. A. (2017). Developing & using interaction geography in a museum. *International Journal of Computer-Supported Collaborative Learning*, 12, 377-399.
- Silvis, D., Taylor, K.H., Stevens, R. (2018a). Community technology mapping: Inscribing places when ‘everything is on the move.’ *International Journal of Computer-supported Collaborative Learning*, 13(2), 137-166.
- Silvis, D., Taylor, K.H. & Stevens, R. (2018b). Reassembling home-work: Mixing “newer” and “older” technologies in home learning environments. In J. Kay & R. Luckin (Eds.), *Proceedings of International Conference of the Learning Sciences (CLS) 2018*. London, England: International Society of the Learning Sciences.
- Star, S.L. (1990). Power, technology, and the phenomenology of conventions: On being allergic to onions. *The Sociological Review*, 38(1), 26-56.
- Stevens, R. (2000). Divisions of labor in school and in the workplace: Comparing computer and paper-supported activities across settings. *The Journal of the Learning Sciences*, 9(4), 373-401.
- Stevens, R. (2010). Learning as a members phenomenon: Toward an ethnographically adequate science of learning. *National Society for the Study of Education*, 109:1, p. 82-97.
- Takeuchi, L., & Stevens, R. (2011). *The new co-viewing: Designing for learning through joint media engagement*. A report of The Joan Ganz Cooney Center at Sesame Workshop and LIFE Center.
- Taylor, K.H., Silvis, D. & Stevens, R. (2017). Collecting and connecting: Intergenerational learning with digital media. In E. Gee, L.M. Takeuchi & E. Wartella (Eds.), *Children and Families in the Digital Age: Learning Together in a Media Saturated Culture*. NY: Routledge.
- Taylor, K.H., Takeuchi, L. & Stevens, R. (2017). Mapping the daily media round: Novel methods for understanding families’ mobile technology use. *Learning, Media and Technology*, 1-15.
- Tulbert, E. and Goodwin, M.H. (2011). Choreographies of attention: Multimodality in a routine family activity. In J. Streeck, C. Goodwin, & C. LeBaron (Eds), *Embodied Interaction: Language and Body in the Material World*. Cambridge, UK: Cambridge University Press.
- Wood, E., Zivcakova, L., Gentile, P., Archer, K., De Pasquale, D. & Nosko, A. (2012). Examining the impact of off-task multi-tasking with technology on real-time classroom learning. *Computers & Education*, 58, 365-374.

# Designing Representations in Deeply Disciplinary Educational Games

Ravit Golan Duncan, Rutgers University, [ravit.duncan@gse.rutgers.edu](mailto:ravit.duncan@gse.rutgers.edu)  
Frieda Reichsman, The Concord Consortium, [freichsman@concord.org](mailto:freichsman@concord.org)  
Kiley McElroy-Brown, The Concord Consortium, [kmcelroy-brown@concord.org](mailto:kmcelroy-brown@concord.org)  
Veronica Cavera, Rutgers University, [veronica.cavera@gse.rutgers.edu](mailto:veronica.cavera@gse.rutgers.edu)

**Abstract:** Educational games are common in classrooms and have been extensively studied in multiple domains. The *Geniventure* game was developed to support student learning of core concepts in genetics via challenges that engage students with genetic phenomena at the molecular level. A core feature of the game is that it simulates behavior of molecular entities (genes, proteins, organelles) with high disciplinary fidelity to how these mechanisms really operate in the cell. In this sense, it is deeply disciplinary. The commitment to disciplinary fidelity presents design challenges regarding the ways in which entities, activities, and mechanisms are represented and manipulated in the game across the biological organization levels. We discuss five distinct types of design challenges that we identified based on data from focus groups with students who played the game and provide design heuristics for addressing these challenges.

## Introduction

The development and research on educational games has increased substantially in the past decade (for recent reviews see Clark, Tanner-Smith, & Killingsworth, 2016; Ravysse, Blignaut, Leendertz, & Woolner, 2017). This work has identified key features that make games more engaging and productive for learning such as having them be pleasantly frustrating, involving well-ordered problems, and providing the player with agency, to name a few (Gee, 2005). This research has also provided some guidelines for the design of successful educational games (e.g. Kafai, 2006; Winn, 2009).

In addition to these critical aspects of game design, we wish to draw attention to considerations of designing *deeply disciplinary* games and in particular designing representations of key disciplinary entities and processes in these games. We use the term *deeply disciplinary* to refer to games that aim to engage students with key disciplinary phenomena and in which the game play involves manipulating entities and mechanisms within these phenomena that may be unfamiliar to students. This construct draws on the notion of *conceptually integrated* games, which integrate domain concepts and relationships into the core mechanics and representations of the game (e.g. Habgood & Ainsworth, 2011) and expands it to include a stronger emphasis on the representation of domain-specific entities and mechanisms in ways that maintain high fidelity to the structure, function, and interactions of these entities. For example, the game *Geniventure* engages students in the genetic study of drakes, the model organism for dragons; students directly manipulate chromosomes, genes, and proteins in order to change the drakes' traits. The representations of these genetic entities in the game are therefore a central component of making the phenomena accessible on the one hand, while maintaining fidelity to the disciplinary entities represented on the other. The tension between accessibility of the representations and their disciplinary fidelity generates interesting challenges in terms of design decisions regarding the representations and the linking of representations across the multiple user interfaces of the game.

Research on students' understanding of disciplinary inscriptions (such as graphs, models, equations) provides some insights that can help inform these design decisions. For example, zoom-in features, alternative perspectives, and control of speed of animations/simulations can help facilitate perception and comprehension of important disciplinary entities and processes (Tversky, Bauer Morrison, & Betancour, 2002). Similarly, tighter coordination and coherence between multiple representations (especially when visualizing phenomena at the macro and molecular levels) is also important to reducing cognitive load and helping students make appropriate connections (Kozma, 2003). In addition, it is often difficult to provide, within the representations, all the needed conceptual and social resources that can help convey the utility and purpose of disciplinary representations/inscriptions (Roth & Bowen, 1999). However, much of this research was conducted in the context of students using and creating inscriptions within, mostly traditional, instructional pedagogies such as the use of inscriptions, animations, or simulations to explain a science concept. Moreover, the research on using computerized models for understanding molecular-level representations and ideas predominantly comes from chemistry (Barnea & Dori, 1996; Williamson & Abraham, 1995; Wu et al., 2000), whereas only a few studies deal specifically with molecular biology (e.g. Pallant & Tinker, 2004; Tsui & Treagust, 2004). Therefore, while



relevant, this body of work does not address the specific challenges associated with using disciplinary representations in game contexts, which are not as conducive to traditional ways of scaffolding and providing information. In such environments, engagement with disciplinary representations and phenomena is in the context of a game challenge and students figure out the key ideas through game play. Exiting the game play to “instruct” or explain the representations or processes can be distracting and problematic in terms of motivation.

Here we present findings regarding middle school students’ understandings of representations and processes in a genetics game and their implications for the design of representations in deeply disciplinary games. The online game, *Geniventure*, is a fictional and narrative-based game in which students are enlisted to help breed dragons with traits that can help them in the struggle against a nearby kingdom’s attacks. As the game opens, students find themselves inside the Mission Control room of a secret underground headquarters for dragon development and breeding. A diverse cast of characters presents students with a series of challenges that address a need for specific traits and an understanding of how those traits are achieved via genetic instructions. Students “travel” to different rooms in the underground compound to breed drakes, manipulate genes, or zoom into cell simulations to interact directly with proteins and DNA. Student actions in a challenge are tracked and upon completion of a challenge students are awarded “crystals” based on how well they performed the task.

The game is designed to help students develop core understandings in genetics including patterns of inheritance (alleles on paired chromosomes control each trait), how random assortment of chromosomes into sex cells results in the observed probabilities of the expression of traits, and how genes on those chromosomes code for proteins that bring about those traits. In this article, we focus predominantly on the game challenges involving proteins (described in more detail below). We conducted focus group interviews with three groups of students who played the protein challenges with the aim of finding out how they understood the nature and role of proteins in mediating the genetic traits involved in the challenges. Our research question is: how are students interpreting and responding to the different representations of molecular entities, mechanisms, and levels of the phenomena (genes, protein, cell, trait) embodied in the game? We next describe the online game, *Geniventure*, and some of the key design decisions that were made in developing the representations of the phenomena.

## The Geniventure game: Protein challenges

The protein challenges are designed to engage students with the mechanisms that connect genes and traits. Genes are DNA sequences that specify instructions to the cell for building protein molecules. Proteins are the workhorses of the cell; they have many different functions, and their functions are closely tied to their structure. Seemingly small differences in the DNA sequences of genes can have profound effects on an encoded protein’s structure and in turn on its function. Different versions of a gene, called alleles, produce different traits by virtue of these differences in the structure and function of the proteins they encode. Thus, proteins are an essential link for productive reasoning about how genes confer traits on individuals.

Proteins are often characterized by experts as molecular machines (e.g., Goodsell, 2009). This is a useful analogy since students understand that machines have functions, and that the form or shape of a machine is essential to its function. In designing the look and feel of proteins in the game, we chose to emphasize their machine-like qualities by using simple machine structures like gears, and we designed their shapes to match



Figure 1. The size challenge.

very obviously with their function. We used color gradients and softened edges to convey their biological nature; influenced by Goodsell’s (2009) use of color in illustrating actual protein structures.

The protein mechanisms we targeted in the game are situated in drake scale cells (analogous to human and other animal’s skin cells) and focus on the production and transport of melanin. We chose drake color as a trait for our protein focus because it is a readily discernable trait and it is constantly and actively maintained by cells through protein-based mechanisms. In the game, students are exposed to the proteins “doing their jobs” and then are given the opportunity to either interfere with what the proteins are doing or to help them along the way, all with the goal of changing the color that the cell produces. The melanosome organelle stores melanin and changes in color based on how much melanin is present. In the game, there are two types of challenges that involve proteins. In the first, which we call the “size” challenge, students try to create smaller or bigger melanosomes. We have reduced the complexity of the mechanism to consist of two proteins, both of which are active in real organisms. One is an enzyme that synthesizes melanin, represented in the game by a set of gears that can assemble melanin, and its “helper protein” that provides stability to the gear enzyme, represented by a shaft that holds the gears together (see Figure 1). The proteins build melanin, which is represented as stars, by assembling the triangular building blocks (white triangles) into melanin stars. Melanosomes change color from orange to gray depending on how many melanin stars are in them; bigger melanosomes are darker gray and smaller melanosomes appear orange. Students work through several instantiations of the size challenge in which they have to either reduce the size of the melanosome (to make an orange drake) by breaking the melanin stars and slowing down the gears, or increase the size of the melanosomes (to make a gray drake) by helping to assemble the gear-shaft complex so it works faster to make melanin stars and thereby grow the melanosome.

In the second challenge, which we call the “gates” challenge, students try to create shiny versus dull/matte drakes (that can be either orange or gray). Shininess (or sheen) is a function of not having any color (no melanosomes) in the outer layer of scale cells, causing those cells to better reflect light and making the scales look shiny. Melanosomes can travel to the outer scale layer through gated channels. These channels are plugged by a protein that selectively lets melanosomes through. The protein is represented by a corkscrew-like structure with a pink tip that can sense the incoming melanosome (see Figure 2). In shiny drakes, the protein plug is not functioning properly and does not ever open the channel; therefore, no melanosomes can get to the outer layer and the scales are shiny. Making the outer layer of scales actually look shiny in the game was a graphics-design problem, and there did not seem to be a simple way of showing sheen. Therefore, we opted for a different solution—having a *sheen indicator bar* positioned at the top of the field of play, within the outer scale layer. When the indicator is “full” it means there are lots of melanosomes in the outer layer and the scale is non-reflective (matte); when the indicator is empty, the scale is shiny. In this challenge, students are tasked with changing drakes from shiny to matte or vice versa by plugging or unplugging the channels.



D = Sheen indicator bar E = Protein plug

Figure 2. The gates challenge.

Both challenges involve students trying to change the color or sheen of the drakes from a provided initial state to a required target state. To help students keep track of their progress (the extent to which the drake is changing) we developed a Heads Up Display (HUD) that includes three elements: a) the initial (start) state of

the drake shown as both a thumbnail of the drake and a thumbnail of a cross-sectional view of the skin tissue; b) the current state of the drake with a thumbnail of the drake only, that actually updates based on progress in the game; and c) the target state shown as dual thumbnails of the drake and cross section of skin. The HUD appears along the right side of the field of play (see Figure 1 & 2) and is present in both challenges.

## Methods

### Study context

The game was played by  $n=51$  middle school students in the context of an eight-week summer program hosted by a community center in a metropolitan city in the North East. The program was open to all middle school students in the city yet the majority of the participating students attended the local public middle school where the program was held. This 6<sup>th</sup>–8<sup>th</sup> grade middle school is ethnically diverse: 38% Black, 25% White; 16% Hispanic, 10%, Asian; with 43.4% economically disadvantaged students. The genetics program in which the game was used ran for 90-minute sessions twice per week and was led and instructed by the Middle School Program Leader and STEM Specialist of the community center. One or two researchers were also present during the genetics programming to assist the instructor. Each session was structured to include some hands-on instructional activities in which students explored various aspects of genetics—for example, students built catapults using normal and ‘mutated’ instructions to see how changes to the instructions can result in changes to the catapults shape and function. These activities were often followed by whole class discussions about the biological meaning of the activity and how it relates to genetics. Students typically spent about 30 minutes, on average, playing the online game during these sessions. The protein challenges, which are the focus of this article, occurred in weeks 3–5 of the summer program. We wish to note that due to various logistical challenges, frequent absences of students, and varying level of engagement of students, the instructor and research team were not able to implement the planned curriculum with high fidelity. Thus, some of the activities and resources that were expected to support student understanding during game play were not fully implemented. While this was a general problem for the ongoing research, we feel that the information from the focus groups is still highly valuable in highlighting issues with the game; some of which may be ameliorated through curriculum activities while others likely require revisions to the actual game design.

### Data collection and analysis

The focus group interviews occurred in week 4 of the program. Student groups were formed based on the game challenges that the students had completed. Focus groups of students who did not reach the protein challenges were not included in this analysis. A total of 9 students participated in the three focus groups included in our analysis. The focus group interview lasted between 19–26 min and involved showing students screenshots of interfaces in the protein game and asking them questions about specific entities or processes occurring in those interfaces such as: “how close is the person playing the game to winning?” “what would you do to make the dragon darker?” “what would you click on next?”. Students took turns answering each question and the interviewer made sure none of the students had anything else to add to the discussion before moving on to the next question. Focus group interviews were transcribed verbatim and analyzed by the research team (authors).

Analysis involved viewing the videotapes together and identifying episodes of interest (e.g. students misunderstanding a representation, students explicitly expressing confusion, etc.). While we cannot claim to have identified all potential problems with the game, and in some cases only one student in the group struggled with a particular issue, we can assume that if these issues happened once, they may happen again. Therefore, we describe the issues that came up without making claims about their prevalence among student players. We grouped the issues we identified under five themes that we see as being potentially relevant for game design in other contexts and discuss them in the results section. We then offer some heuristics for designing representations for deeply disciplinary games that could help address the general problems we identified.

## Results

Before we discuss specific themes of problems with the game we wish to note that overall the students did seem to interpret representations, game phenomena, and game play as intended by the designers. Many of the features that we had hoped they would notice about the representations of genes, proteins, and traits in the game were indeed attended to by the students. It was also clear that students who had more prior knowledge in genetics, which came through in the interviews without us explicitly asking questions about genetic concepts, understood the game phenomena, mechanisms, and entities better than students who did not. This is not surprising and has been documented extensively in prior research (Cook, Wiebe, & Carter, 2007; Kindfield, 1994).

The five thematic problems we identified are: (1) incongruence between phenomenological knowledge and representations, (2) unintended distractors, (3) non-salient ontological distinctions, (4) misinterpretation of linked representations, and (5) confusing game progress and disciplinary process indicators. We discuss them in turn, by first describing the problem as it arose in the focus groups, then providing our interpretation of what is the specific problem is a case of, and lastly discussing some possible solutions for this problem in the game.

### **Theme I: Incongruence between phenomenological knowledge and representations**

As we noted earlier, the building blocks of melanin (white triangles) can be assembled into melanin “stars”, which are initially light orange. As they accumulate, however, they become darker—as does the overall melanosome (Figure 1). In one of the initial challenges, students are tasked with making a darker gray drake by interacting with the process of melanin star production in the melanosomes of a lighter orange drake. One of the students in the focus groups noted a confusion he initially had when playing this challenge. He assumed that *breaking* the bright orange stars would lead to darker melanosomes. We interpret this situation as the student drawing on phenomenological primitive-like knowledge (diSessa, 1998) of the sort- “less bright means dark” and acting on it to break apart (remove) the bright-looking stars. While the student eventually recognized that the strategy of breaking stars had the opposite effect (smaller melanosome), the student’s initial interpretation of the representation (bright stars) and its relation to the phenomenon (skin color of drake) is important. This is because the student’s initial reaction was intuitive and yet counterproductive given the specific representations used. We argue that in general terms this is a problem of having representations and/or processes that behave in ways that counter, or are incongruent with, students’ phenomenological knowledge. By this we do not mean that the representations are complex or unfamiliar, or that the processes are counterintuitive. Rather we mean that the actual choice of color (or other feature) of the representation, from a semiotic perspective, is incongruent with students’ knowledge of how entities behave in the world at a basic phenomenological level. The decision to make the stars light orange on a dark blue background was driven by the aesthetics of having the stars stand out as brighter objects on the screen. In hindsight, this was problematic in that it cued students to take an intuitive but unproductive action. Fortunately, the problem also has a fairly simple solution of changing the background color of the field of play to a lighter color such that the stars are not so bright that they appear to generate light.

### **Theme II: Unintended distractors**

Research has shown that students, especially those with low prior knowledge, tend to select the most noticeable features of the representations for further processing and may ascribe meaning to un-important features of representations when mapping between macro and molecular representations (Cook, Weibe, & Carter, 2007; Seufert, 2003). While we found that students did notice most of the relevant details of entities and processes in the game, they also noticed and attempted to interpret details that were not relevant. For instance, in the melanosome size challenges, when a melanosome is small the triangular building blocks of melanin tend to cluster in a higher concentration around the melanosome; when the melanosome is larger the same number of triangles are still there but seem less concentrated due to the larger circumference of the melanosome. This is a randomly occurring situation that is an artifact and not a feature of the representation. However, one student assumed that the discernable clustering of triangles around the melanosomes as somehow contributing to the change in melanosome color. In general terms, this is a problem of unintended (unforeseen) and sometimes unavoidable distractors in a complex disciplinary display of mechanisms. The clustering of triangles, while not a designed behavior per se, is nonetheless reflecting the real-world complexity of the mechanisms and the random nature of particle movement in the cell. A solution therefore cannot be to avoid this occurrence. The point we wish to make is that student will attend to all features of the representation, those that were deliberately engineered as well as those that were not (and are meaningless). Our solution here will be to try and alter the underlying dynamics of the representation such that the clustering behavior is less frequent.

### **Theme III: Non-salient ontological distinctions**

In the game, there are several distinct kinds of entities such as proteins, their substrates (triangles) and their products (stars). As noted above, we used representational elements that resemble little machines to convey the functional aspect of proteins (as molecular machines that carry out functions in the cell). In contrast, we used simpler geometrical shapes to represent protein substrates and products. Our attempts to convey this ontological distinction through semiotic cues were not always interpreted as intended by students. While students did notice the more complex and gear-like shapes of proteins, they did not view these as inherently different kinds of things from the stars and triangles, and did not realize the biological significance of the differences. One might argue that this is not surprising given that students lacked prior knowledge in the domain and therefore were unlikely to understand the biological meaning of the differences in representations. We agree that students’

domain knowledge plays a role here, however, if the game is intended to help novices learn about the biological significance of key entities and mechanisms in genetics then finding ways to convey important ontological distinctions in the discipline is an important user interface design goal. It is not clear to us how we can necessarily improve the representations in order to make the meaning of the distinctions more salient. The solution to this problem will likely involve adding information to the game through hints or other cues that help students make sense of the differences they are already noticing in the representations of entities. Or it may be that the sense making needs to occur in the context of group/class discussion in which the teacher facilitates student reflection on the differences and what they mean.

#### **Theme IV: Misinterpretation of linked representations (HUD)**

A central representational feature in the protein challenges is the HUD, which provides real-time information about the current color of the drake relative to the initial state of the challenge and the target state. The HUD provides representations both at the cell level, using a cross sectional skin tissue view, and at the whole organism level by displaying a thumbnail of the drake (see Figure 1). The “current state” cell and drake representations in the HUD are linked to the field of play. We identified two representational problems with the HUD design. First, some students did not understand that the centered part of the HUD was representing current state. They instead thought it was representing an intermediate state between the initial and target states and therefore when shown a screenshot including the HUD, they struggled to determine how far along the player was on that particular challenge. This type of confusion of the HUD was surprising to us given the central role the HUD plays in helping the players determine their progress in the challenge.

The second problem with the HUD related to the specific representations chosen to portray the cell/tissue levels, and was more severe for the “gates” challenges than for the “size” challenges. In part, this relates to the complexity of the phenomenon in each challenge. The gates challenges entail understanding that melanosomes move (or are blocked from moving) from one layer of cells in the skin into another, more surface layer. The ideas that skin is multilayered and that organelles like melanosomes can travel between cells were likely new to students, and while they could successfully play the challenge it was clear that many did not understand the underlying biological phenomenon. The HUD in this case was particularly confusing because it provided a cross section of the two layers of cells. Students were able, for the most part, to interpret their state in the game using the HUD but could not explain what the representations in the HUD were actually showing (i.e. where is this in the drake?). We believe that students may be unfamiliar with cross sectional images and how to relate them to a “view from above” perspective. While they could pattern-match entities in the field of play to their smaller representations in the HUD, they did not understand how the zoomed-out view of the HUD is spatially oriented in relation to the representations on the field of play. Stated in general terms this is a problem of both misinterpreting the linked representations and being unfamiliar with particular disciplinary inscriptions. A solution to this problem could be to change the representation used in the HUD from a cross section view to some other more familiar view. However, the cross-sectional view is a fairly prevalent disciplinary way of showing locations of cells within a larger organismal context and there are reasons to keep it given the work it can do (when someone understands it). If the representation remains, the solution to this problem will likely entail providing more cues about the relationships between the spatial orientation of the HUD and the field of play. Perhaps adding a short video that shows how you get from one perspective to another (at the start of the gate challenges) will allow students to “see” the connections between these linked representations.

#### **Theme V: Confusing game progress and disciplinary process indicators**

In the gates challenges, students try to change the sheen of the scales of the dragon from matte to shiny or vice versa. As noted earlier figuring out how to represent an increase or decrease in scale sheen in the game was a design challenge that we opted to solve by incorporating an indicator bar for sheen that shows how shiny the outer cell layer is at any point in time (an empty bar means the scales are shiny and a full bar means the scales are matte). The sheen indicator bar sits within the field of play (see Figure 2) and includes a small indicator arrow that shows the target state (empty or full bar). The location of the sheen indicator bar compounded with the complexity of the mechanism of the phenomena (students struggled to understand why it is that having no melanosomes in the outer cell layer makes a dragon shiny) resulted in some confusion about what that bar was actually representing. Some students understood it as showing sheen (as intended), while others thought it showed progress in the challenge (like the HUD). In the latter case, this was a problem when the target dragon was shiny and the goal was to have an empty bar. The discrepancy between the bar being empty and the challenge being successfully completed was confusing to some students who expected a full bar to indicate the challenge was complete. In general terms, the problem here is one of inconsistent use of general game navigation cues (like a progress bar) to represent mechanistic processes of a disciplinary phenomenon. Overall

using a progress-bar type representation to show changes in a mechanism is not an issue in and of itself; in fact, the HUD provides such support in the protein challenges. The case here is that this was a new navigational/indicator feature that was used sporadically in *only* one set of challenges- the protein gates challenges. One possible solution, therefore, is to either include a similar indicator bar in the other challenges or to try again to figure out how to represent sheen within the phenomenon itself (in the outer layer of cells).

## Discussion

The focus group interviews, while limited in scope, nonetheless provided us with valuable information about the ways in which students attended to, interpreted, and acted on the representations of phenomena in the game. The problems we identified reflected design and usability challenges associated with representing entities and mechanisms that are inherent and endemic to the discipline of genetics. In this section, we wish to ‘pop up a level’ and discuss some of the implications from this work more generally and provide a set of heuristics that can guide the design of representations in deeply disciplinary games.

The first heuristic, *balancing aesthetics, disciplinary fidelity, and phenomenological congruence*, stems from themes I and II. Deeply disciplinary games need to maintain high fidelity to the discipline in terms of how entities, activities, processes, and so on are represented. They also need to engage and attract students, and whenever possible they should be consistent with phenomena and behaviors that students may be familiar with (e.g. correspond p-prims that students likely hold). Balancing these three demands is not trivial and decisions that privilege aesthetics over fidelity or congruence (intentionally or not) could be costly in that it may take students longer to figure out the game dynamics. In addition, while disciplinary fidelity may seem like the most important consideration, adhering to it may have inadvertent consequences if it leads to meaningless aspects of a mechanism or entity being overly salient (e.g. clustering of triangles around the melanosome, Figure 1). It may be prudent at times to forgo some of the fidelity in favor of avoiding such situations.

The second heuristic, *pointing out core disciplinary distinctions*, stems from theme III. We have found that students were able to notice differences between ontologically distinct entities, but they did not recognize the biological significance of these differences. Altering the representations to make them even more different from each other, and different in ways that better reflect their ontological origins, may minimize the problem to some extent. Yet such changes are unlikely to be enough and we believe that additional scaffolds are needed to help students grasp the disciplinary nature and significance of the differences. The scaffolding design framework developed by Quintana et al., (2004) included a guideline about *making disciplinary strategies explicit in learners’ interactions with the tool* (p. 345), which addresses the similar problem of students not having disciplinary knowledge to guide them in reasoning about problems in the discipline. The heuristic we provide builds on and extends this guideline by suggesting that one also needs to make explicit *ontological distinctions and features* of core entities and mechanisms. For example, in our game, it will be beneficial to explicitly point out that proteins resemble little machines because they act as such in the cell, whereas the melanin stars are inactive molecules comprised of smaller building blocks. Such “pointing out” needs to be done in the context of the game in ways that do not interrupt the game play. In a sense this conceptually important ontological distinction was not integrated well enough in the game. This presents a true design challenge in terms of making these key distinctions better conceptually integrated into the game. Another potential solution may be to draw on the Constructed Authentic Representations (CAR) principle (Holbert & Wilensky, 2014) and to allow students to build (through manipulation of DNA sequences- the instructions) proteins with particular functional domains. This may make more salient the distinction between what can be constructed (protein machines) by changing the genetic instructions and what is synthesized by using the protein machines.

The third, and final, heuristic, *supporting the linking of phenomena states to progress indicators*, stems from themes IV and V. We found that students had difficulties in relating both the HUD and the sheen indicator bar to the state of affairs in the field of play. There are actually multiple issues that contributed to this problem including lack of disciplinary knowledge of common representations (cross sectional views of tissues), the difficulty in representing sheen, and misinterpreting the indicator bar to be a general progress indicator. Therefore, in designing HUDs one needs to consider what types of links to the phenomenon are represented. Links could be temporal, showing ongoing progress of a mechanism as in the start-current-target states in the HUD; and/or the link can be spatial, showing where the mechanism is occurring in relation to the rest of the phenomenon as in the cross-sectional images of the cell and the images of the drakes in the HUD. In both cases students need support in making the connections and understanding what the HUD is showing. It is likely that the HUD in the protein challenges by itself is insufficient to support this linking and that additional scaffolds may be needed to help orient students to relevant relationships (such as an animation that connects the HUD and the phenomenon).



In designing *Geniventure* to accurately reflect complex genetic phenomena at the molecular level, we made many design decisions regarding how to represent key disciplinary entities and processes in the game in ways that can convey their biological characteristics. In many cases, the decisions proved effective and students were able to comprehend what they are doing in the context of the game and make progress across challenges. In other cases, the decisions hindered student progress. Moreover, even students who were able to progress substantially in the game did not always understand some of the underlying biology, such as the differences between proteins and other molecules in the cell (e.g. melanin). This underscores the importance of providing supporting curriculum and instruction to help students reflect on and generalize the mechanisms they are seeing in the game and their biological significance. A key question that arises is how can we decide on the division of scaffolding labor between the game, curriculum, and instruction (teacher and peers). This issue is even thornier in educational game contexts where there is a risk of making the game feel more school-like in ways that disengage students.

## References

- Barnea, N., and Dori, Y. J. (1996). Computerized molecular modeling as a tool to improve chemistry teaching. *Journal of Chemical Information and Computer Sciences*, 36: 629–636.
- Clark, D. B., Tanner-Smith, E., & Killingsworth, S. (2016). Digital games, design, and learning: A systematic review and meta-analysis. *Review of Educational Research*, 86(1), 79–122.
- Cook, M., Wiebe, E. N., & Carter, G. (2007). The influence of prior knowledge on viewing and interpreting graphics with macroscopic and molecular representations. *Science Education*, 92, 848–867.
- diSessa, A. (1988). Knowledge in pieces. In G. Forman & P. Putall (Eds.), *Constructivism in the computer age* (pp. 49-70). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Ge, J. P. (2005). Learning by Design: Good video games as learning machines. *E-Learning*, 2(1), 5–16
- Goodsell, D. S. (2009) *The Machinery of Life*, 2nd ed., Springer, New York.
- Habgood, M. P. J., & Ainsworth, S. E. (2011). Motivating children to learn effectively: Exploring the value of intrinsic integration in educational games. *Journal of the Learning Sciences*, 20(2), 169–206.
- Holbert, N., & Wilensky, U. (2014). Constructible authentic representations: Designing video games that enable players to utilize knowledge developed in-game to reason about science. *Technology, Knowledge and Learning*, 19(1–2), 53–79.
- Kafai, Y. B. (2006). Playing and making games for learning: Instructionist and constructionist perspectives for game studies. *Games and Culture*, 1(1), 36–40.
- Kindfield, A. C. (1994). Biology Diagrams: Tools to Think With. *Journal of the Learning Sciences*, 3(1), 1–36.
- Kozma, R. (2003). The material features of multiple representations and their cognitive and social affordances for science understanding. *Learning and Instruction*, 13(2), 205–226.
- Pallant, A., and Tinker, R. (2004). Reasoning with atomic-scale molecular dynamic models. *Journal of Science Education and Technology*, 13(1), 51–66.
- Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J., Fretz, E., Duncan, R. G., et al. (2004). A scaffolding design framework for software to support science inquiry. *Journal of the Learning Sciences*, 13(3), 337–386.
- Ravyse, W. S., Blignaut, A. S., Leendertz, V., & Woolner, A. (2017). Success factors for serious games to enhance learning: A systematic review. *Virtual Reality*, 21, 31–58.
- Roth, W., & Bowen, G. (1999). Complexities of graphical representations during ecology lectures: an analysis rooted in semiotics and hermeneutic phenomenology. *Learning and Instruction*, 9(3), 235–255.
- Seufert, T. (2003). Supporting coherence formation in learning from multiple representations. *Learning and Instruction*, 13(2), 227 – 237.
- Tsui, C.Y., & Treagust, D. (2004). Motivational aspects of learning genetics with interactive multimedia. *The American Biology Teacher*, 66, 277–285.
- Tversky, B., Bauer Morrison, J., & Betrancourt, M. (2002). Animation: Can it facilitate? *International Journal of Human-Computer Studies*, 57, 247–262.
- Williamson, V.M., & Abraham, M.R. (1995). The effects of computer animation on the particulate mental models of college chemistry students. *Journal of Research in Science Teaching*, 32, 522–534.
- Winn, B. M. (2009). The Design, Play, and Experience Framework. In *Handbook of Research on Effective Electronic Gaming in Education* (Vol. 3, pp. 1010-1024). Hershey, PA: IGI Global.
- Wu, H.K., Krajcik, J.S., & Soloway, E. (2000). Promoting understanding of chemical representations: Students' use of a visualization tool in the classroom. *Journal of Research in Science Teaching*, 38, 821–842.

## Acknowledgments

The research reported was funded by the National Science Foundation under Grant No. DRL-1513086.

# The Knowledge Building Network Pilot Project: An Exploration of Emergent Designs to Enhance Collective Teacher Efficacy

Leanne Ma, OISE/University of Toronto, [leanne.ma@mail.utoronto.ca](mailto:leanne.ma@mail.utoronto.ca)  
Monica Resendes, OISE/University of Toronto, [monicaresendes@gmail.com](mailto:monicaresendes@gmail.com)  
Marlene Scardamalia, OISE/University of Toronto, [marlene.scardamalia@utoronto.ca](mailto:marlene.scardamalia@utoronto.ca)  
Karen Dobbie, Ontario Ministry of Education, [karen.dobbie@ontario.ca](mailto:karen.dobbie@ontario.ca)

**Abstract:** Ontario's Leading Student Achievement Project aims to foster professional collaboration among principals, vice-principals, and teachers in order to build capacity for improving student achievement and well-being. During the 2017-2018 school year, a multi-level, multi-district KB network was initiated to spread Knowledge Building in elementary schools across the province. This study traces the evolution of KB and KF practices in relation to changes in teacher efficacy and positive student outcomes. Over the span of a school year, teachers co-designed principle-based practices with their students to foster collective discourse, idea improvement, community norms, and meta-cognition/meta-talk in their classrooms. Results indicate that the initiative was a success, with the majority of educators reporting professional growth through collaboration, as well as willingness to lead their own KB network the next year.

## Introduction

Teacher efficacy is defined as “teachers’ perceptions of their own ability to bring about desired student outcomes” (Takahashi, 2011). Research has consistently demonstrated a strong relationship between teacher efficacy, classroom practices, and student achievement (e.g., Hattie, 2012; Goddard, Hoy, & Woolfolk Hoy, 2000). Teachers with high efficacy set more challenging goals, show greater effort and persistence, and are more willing to try new teaching approaches (Tschannen-Moran & Woolfolk Hoy, 2001). Studies also indicate that teachers’ individual efficacy is interlinked with collective efficacy (Schunk & DiBenedetto, 2006), where collective teacher efficacy refers to the collective perception that “teachers in a given school make an educational difference to their students” (Tschannen-Moran & Barr, 2004). Enabling conditions of collective teacher efficacy include shared goals (Kurz & Knight, 2003), group mastery experiences (Goddard, 2001), and opportunities for teacher leadership (Derrington & Angelle, 2013). Thus, school administrators (i.e., principals and vice-principals) play a vital role in fostering collective teacher efficacy (Ross & Gray, 2006; Leithwood & Jantzi, 2008). Taken together, this line of work suggests that teacher efficacy, both at the individual and collective levels, are important for enhancing teaching practices and student achievement.

Over the past 13 years, the Ontario Principals’ Associations (1), in partnership with the Ministry of Education, have been fostering professional collaboration among principals, vice-principals, and teachers in order to build their collective capacity for improving student achievement through the Leading Student Achievement: Networks for Learning project (LSA). The project is guided by the evolving “LSA Theory of Action” (Leithwood, 2018), which highlights conditions for school success along the rational path (e.g., Knowledge Building), the emotional path (e.g., collective efficacy), the organizational path (e.g., professional learning communities), and the family path. As such, one key priority for the LSA project is to create professional learning networks for Knowledge Building/knowledge creation. This study explores supports for implementing Knowledge Building at scale by tracing the evolution of classroom practices in relation to changes in teacher efficacy and positive student outcomes.

## The LSA Knowledge Building network pilot project

During the 2017-2018 school year, a pilot project was created for educators to learn, share, co-design, and refine Knowledge Building practices that supported student learning and well-being. 32 teachers and 8 administrators from 17 elementary schools across 8 boards were brought together to create a multi-level, multi-district network to promote the cross-fertilization of ideas and spread of classroom innovations across the province. Each teacher was committed to implementing Knowledge Building (KB) pedagogy and Knowledge Forum (KF) technology in at least one subject area, such as math, science, or social studies. Members of the network met five times between September and May to engage in collaborative design of KB and KF practices. Throughout the year, teachers were provided with customized capacity-building sessions and on-site KF support as often as needed to help advance their design work. Data sources for this study include: teachers’ KB practices presented at bi-monthly



meetings and teachers' responses to end-of-term surveys, including their reflections about the design process and observations of how their practices impacted their students.

### Knowledge Building practices

One of the initial challenges of the pilot project is introducing Knowledge Building (Scardamalia & Bereiter, 2014), which, unlike other pedagogical approaches, adopts a principle-based approach to classroom design. Many teachers found the conceptual complexity of the 12 KB principles (Scardamalia, 2002) to be less than intuitive and struggled with translating them into practice. To help scaffold the development of their understanding, teachers were encouraged to study the KB Gallery (Resendes & Dobbie, 2017), as well as multimedia resources on <http://thelearningexchange.ca>, including: descriptions of the 12 KB principles, concrete tools and strategies for getting started in the classroom, video series featuring teacher, student, and parent testimonials, and podcast series to provoke reflection and classroom planning. The bi-monthly meetings served as opportunities to discuss their attempts at principle-based classroom design (both successful and failed ones), ask questions and troubleshoot, receive feedback from peers, and plan for their next iteration.

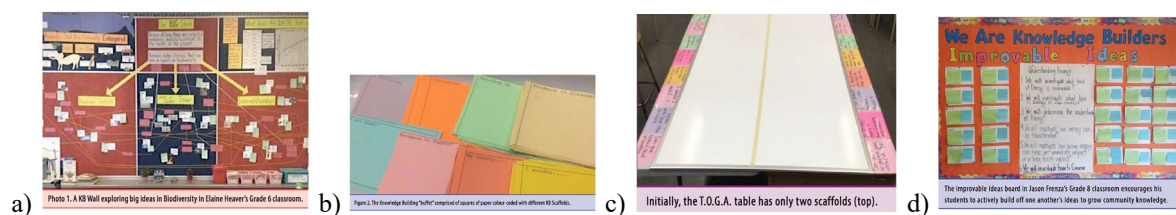


Figure 1. a) KB wall, b) KB scaffolds, c) KB table, and d) Improvable Ideas board.

Figure 1 shows some KB practices invented by Ontario teachers in the companion guide to the KB Gallery, including: a) KB walls, which visualizes connections in the community knowledge; b) KB scaffolds, which promotes diverse contributions to the community discourse; c) KB table, which makes student ideas visible to all members of the community; and d) Improvable Ideas board, which shows design iterations on student ideas. These practices have the shared goal of bringing student ideas to the center of classroom activities and fostering students' sense of collective responsibility for knowledge advancement (Scardamalia, 2002). Put differently, these practices help teachers realize various KB principles in their classroom, such as *community knowledge*, *collective responsibility*, *Knowledge Building discourse*, *democratizing knowledge*, and *improvable ideas*. The year-long challenge for network participants was to adopt these practices into their own classrooms and schools, refine them through deepening engagement with all 12 KB principles, as well as co-design new KB practices with their students.

### KB network meetings

During the KB network meetings, 3 teachers, 3 teaching teams, and 6 school teams presented their classroom designs and re-designs. Each presentation highlighted the implementation of KB practices in areas such as math, science, history, geography, health, and robotics. On average, each classroom design included 4 to 5 practices (Figure 2). The most commonly used practices were: KB scaffolds (9), KF community (9), KF analytic tools (8), KB walls (7) and community norms (6).



Figure 2. KB practices used by elementary network teachers.

It can be seen that neither face-to-face practices (e.g., KB walls, KB circles, KB table) nor online practices (e.g., KF community, KF analytic tools) alone could account for half of KB practices presented, suggesting that most teachers adopted a blended approach to their KB practices. In particular, KB scaffolds and community norms were used in both face-to-face and online contexts, which points to their importance in facilitating the integration of KF technology with KB practices in the classroom.

Given the fact that most teachers implemented multiple KB practices at once, the average classroom design reflected 8 KB principles. The most common KB principles translated into practice were: *idea diversity* (12), *KB discourse* (11), *democratizing knowledge* (11), *improvable ideas* (10), and *real ideas, authentic problems* (9). Whereas presentations at the beginning of the school year mostly focused on *real ideas, idea diversity*, and *KB discourse*, presentations toward the end of the school year focused more on *community knowledge, democratizing knowledge*, and *improvable ideas*. This supports the notion that while it is easy to engage students in idea generation, it is challenging to engage them in idea improvement (Scardamalia & Bereiter, 2003). One common practice that teachers consistently used to foster *improvable ideas* was the integration of KB scaffolds into day-to-day class discussions.

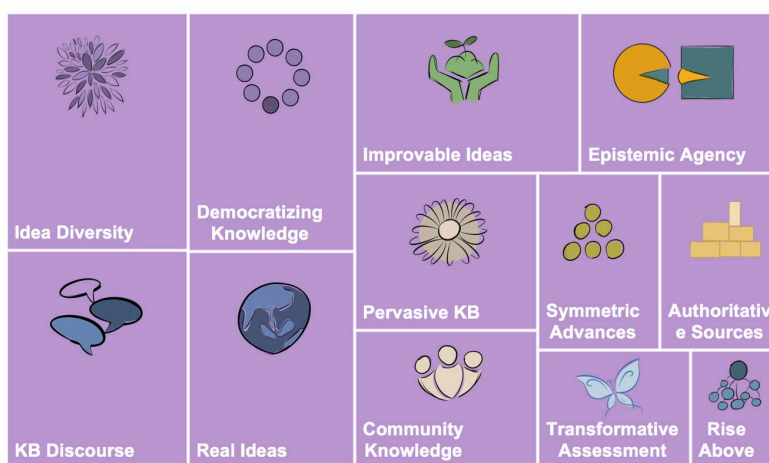


Figure 3. KB principles used by elementary network teachers.

Taken together, it appears that *KB discourse*, which aims to shift engagement norms of idea generation toward idea improvement, is a fundamental principle to fostering and sustaining a KB community regardless of students' grade or subject area.

It is also interesting to note that the least common KB principle in classroom designs was *rise above*, which reinforces the overarching challenge of seeking conceptual coherence among diverse ideas in the community knowledge. Indeed, one math teacher attempted to design for *rise above*. After studying formulae for area and perimeter of 2D shapes, students were encouraged to develop theories for how to calculate the surface area and volume of 3D shapes. The teacher provided students with a *real, authentic problem* of filling the volume of a sinkhole to see whether students could create formulae for rectangular prisms and cylinders. Students were so captivated with the problem that in addition to finding the formula for a cylinder (including seeking *authoritative sources* to learn about circumference of circles), they were creating formulae to calculate the cost of materials and labour to fix the hole! For another science teacher, *rise above* emerged naturally as students were studying biodiversity in science and climate change in social studies. After spending weeks exploring the relationships between predators, preys, and invasive species in various ecosystems, as well as discussing how humans have an impact the environment, students brought their *symmetric knowledge advances* together to develop the *rise above* theory that humans are an invasive species.

It can be seen that complex, interdisciplinary problems give rise to rich discussions in the classroom. Therefore, it comes as no surprise that deeper KB practices (i.e., involving multiple KB principles) are more likely to occur when teachers implement KB across multiple subject areas. Statistical analyses reveal a positive correlation between the number of subject areas a teacher used for KB and the number of KB principles reflected in their designs ( $r = 0.76, p > 0.01$ ), as well as the number of KB practices a teacher used and the number of KB principles reflected in their designs ( $r = 0.74, p > 0.01$ ). This appears to confirm the notion that Knowledge Building requires an enculturation approach to principle-based practice – the more opportunities teachers and students have to engage in sustained, creative work ideas, the deeper they go with KB, eventually unlocking all 12 principles (Scardamalia & Bereiter, 2003; Resendes & Dobbie, 2017).

### KB network surveys

At the end of the fall and winter terms, teachers reflected on their KB practices and completed surveys (27 responses were collected in fall, 25 responses were collected in winter). Using the KB progressions in the KB Gallery resource, teachers rated their progress along four dimensions of KB practice: fostering collective discourse (orange), community norms (yellow), developing ideas (green), and meta-cognition/meta-talk (blue). Figure 4 shows the proportion of teachers in early stages (light colours), developing stages (medium colours), and deepening stages (dark colours) for each dimension of practice. In the fall term, the majority of teachers were in early practice, whereas in the winter term, the majority of teachers were in developing practice. Change scores reveal a 55% decrease in early practices and 68% increase in developing practices between the two terms, suggesting that teachers were engaged in continual development of their KB practices throughout the year.

The survey also provided opportunities for teachers to elaborate on their KB practices in open-ended questions. 20 out of 25 of teachers believed that the key community norm was fostering a sense of safety and community. This includes promoting a growth mindset, openness to different perspectives, respectful communication, celebrating risks, and valuing mistakes. Whereas in the fall term, more teacher-directed practices were reported to scaffold students into KB practices, in the winter term, many teachers were engaging students in co-design of KB practices. For example, in the fall, many teachers created posters and cue cards of KB scaffolds while modelling how to use the them during class discussions. In the winter, they listened for students' use of the KB scaffolds while encouraging them to design new ones to advance the community discourse and expand the class's repertoire of KB scaffolds (e.g., "I wonder", "Another way of thinking...", "We should revise this..."). Some teachers were even beginning to recognize curricular connections, suggesting that they were shifting toward deeper practices. Halfway through winter term, one health teacher realized that implementing KB across the curriculum would have been less effortful on her part and more rewarding for the students than carving out their designated weekly KB time. Recall that in the teacher presentations, deeper KB practices involved implementing KB across multiple subject areas.

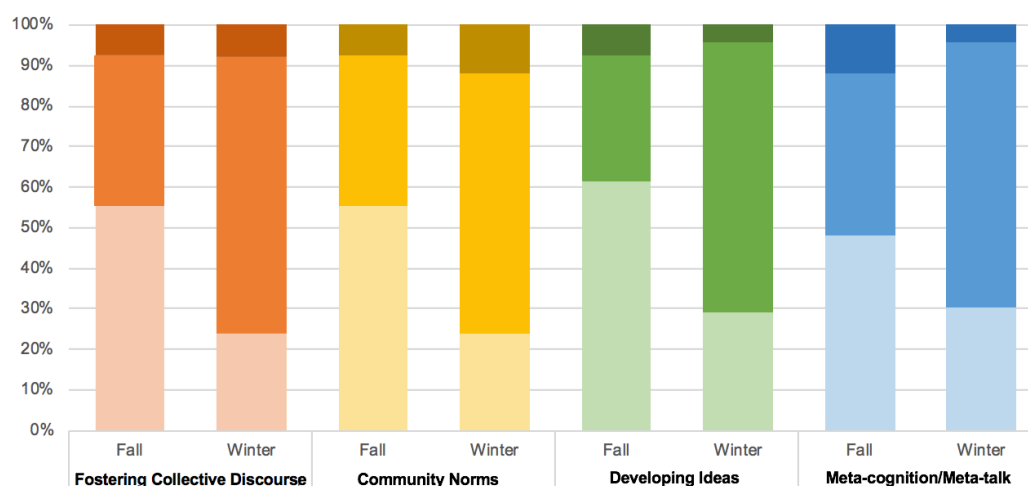


Figure 4. Teacher self-assessments on KB Progressions.

Regarding the most common practices used for fostering community discourse, developing ideas, and meta-cognition/meta-talk, KB scaffolds, KB circles and KB walls were most commonly used in fall, whereas KB scaffolds, KB circles and KF were most commonly used in winter. Given the considerable overlap in responses across the four dimensions, it is likely that the all-in-one nature of KF (i.e., KB scaffolds for collective discourse, endless KB walls for developing ideas, analytic tools for meta-cognition/meta-talk) helped teachers deepen their KB practices. For example, while teachers were encouraging students to use KB scaffolds across both terms, it was not until when students were using the KB scaffolds in KF notes that teachers were able to assess student contributions through the analytic tools. When students were provided with the opportunity to reflect on their contributions through the analytic tools, both the quality and quantity of student engagement was enhanced. Students were excited about the analytic tools and would self-initiate their use to regulate their own contributions. Put differently, the intentional integration of KF into KB practices reflects teachers' growing understanding of the KB principles and their increasing efficacy in balancing the face-to-face and online affordances for advancing community knowledge. Below are a few teachers' reflections on the impact of their KB practices on student learning and well-being:

“Through KB circles students are beginning to build on each other’s ideas while discussing our emotions and ways that we can regulate our emotions.”

“KF allowed students with anxiety to feel like an important, active member of the community by sharing important ideas that others actually read!”

“Students are more engaged – students have a voice and want to share their ideas and communicate their thinking. The students are gaining new tools for their tool belt... and are able to ask questions and reflect on their peers’ process as well as their own learning.”

Overall, responses in the end-of-year survey indicate that both teachers and administrators found the network sessions immensely helpful for: deprivatizing classroom practices; troubleshooting and advancing their designs; hearing from different school boards, roles, and perspectives; exposing them to new ideas and strategies; and opening new opportunities for collaboration. Educators at all levels almost unanimously (all except one who was unsure) expressed interest in continuing participation in the multi-district KB network the following year, with 20 out of 25 willing to start a smaller KB network in their schools and/or school boards. Below are some additional insights and reflections from network participants:

“It is always great to hear ideas from other schools and boards, and different perspectives (teacher, admin, students). It is inspiring to learn with educators from other areas. This session gave me the opportunity to reflect on my classroom and program and consider how to further develop student thinking and deepen practice.”

“Being able to talk to other teachers and having discussions around how to approach some of my challenges. Learning new ways to use KB in my classroom. It validates my pedagogy and has helped to re-invent myself as an educator.”

“I’d like to continue having teachers, students, and administrators share what is happening inside our classrooms and provide time for us to pair up and work together across the regions when working on like ideas.”

“Interesting to hear where other schools and other boards are in this journey. Different schools/boards are strong in different areas and it is great to share our information to help the teaching/learning community as a whole develop and grow together.”

It can be seen that teachers were taking ownership of their classroom practices. Furthermore, the network sessions helped them see their own practices as improvable, and over time, they were assuming collective responsibility for advancing the design work of their community. In other words, members of the KB network demonstrated intentional efforts to raise their individual efficacy and collective efficacy. As an outcome of this pilot project, 16 teachers and administrators attended the KB Summer Institute to share their advances and explore possibilities for the 2018-2019 school year. This experience helped them reframe their practices around less commonly targeted KB principles over the past year, such as *community knowledge*, *pervasive KB*, and *symmetric knowledge advancement*. They are now linking with international KB communities and building their own networks to deepen their practices and grow capacity in their classrooms, schools, and boards. In the words of one teacher:

“My goal is to continue to have my students... participate in this network universally by having them make effective contributions during online discourse (Pervasive KB) on the existing global issues that prevail today. It’s phenomenal observing what solutions/innovations these students come up with that could potentially one day benefit the Public Good!”

## Conclusions and implications

Multi-level networks and partnerships are powerful mechanisms for building capacity as well as spreading Knowledge Building within a system (Laferrriere, et al., 2010; Chan, 2011). After the success of the KB Tri-Board Project (Resendes et al., 2016), Knowledge Building has spread across the province to involve more than 11 districts. The purpose of the 2017-2018 KB Network pilot project was to create opportunities and supports for teacher collaboration and design within and across districts in Ontario. Over the span of one school year, teachers

adapted existing KB practices – such as KB scaffolds, KB walls, KB circles, KB walls, and KF communities – into their classrooms and refined those practices with their students.

In working toward translating the 12 KB principles into meaningful classroom designs, teachers were engaged in the conceptual work of shifting the relations between ideas, students, and themselves in the classroom (Teo, 2014; Toth & Ma, 2018). The more KB principles they integrated into their designs, the deeper they went with their KB practices, the more engaged their students. At the same time, the bi-monthly network sessions allowed teachers to deepen their understanding of the KB principles by discussing concrete practices from other teachers' classrooms and reflecting on the adaptability of their classroom designs. Studies on KB in professional learning communities indicate that collaborative design helps teachers become more open and proactive toward their practices (Vokatis & Zhang, 2016), as well as orient towards more innovative practices (Hong, Chai, & Hung, 2015).

This study suggests that the LSA KB network pilot project successfully provided enabling conditions for teachers to engage in collaborative design as means to improve their KB practices. Consistent with past research in professional development (Hoy, 2000), both direct experiences and vicarious experiences were conducive to enhancing teacher efficacy. Recall that teacher efficacy is key to improving teaching and student learning (Takahashi, 2011) while simultaneously enhancing professional commitment (Ware & Kitsantas, 2007). This study suggests that Knowledge Building goes beyond fostering teacher efficacy to igniting both teachers' and students' epistemic agency. While teacher reports indicate that their practices have a positive effect on students' engagement, learning, and well-being, additional analyses are underway to further understand the effects of sustained participation in dynamic multi-level, multi-district KB networks on student achievement outcomes.

## References

- Chan, C. K. (2011). Bridging research and practice: Implementing and sustaining knowledge building in Hong Kong classrooms. *International Journal of Computer-Supported Collaborative Learning*, 6(2), 147-186.
- Derrington, M. L., & Angelle, P. S. (2013). Teacher leadership and collective efficacy: Connections and links. *International Journal of Teacher Leadership*, 4(1), n1.
- Goddard, R. D. (2001). Collective efficacy: A neglected construct in the study of schools and student achievement. *Journal of Educational Psychology*, 93(3), 467.
- Goddard, R. D., Hoy, W. K., & Hoy, A. W. (2000). Collective teacher efficacy: Its meaning, measure, and impact on student achievement. *American Educational Research Journal*, 37(2), 479-507.
- Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. Routledge.
- Hong, H. Y., Chai, C. S., & Hung, G. T. (2015). Principle-based guidance to foster adaptive teaching practice. In O. Lindwall, P. Hakkinen, T. Koschman, P. Tchounikine, & S. Ludvigsen (Eds.), *Proceedings of the 11<sup>th</sup> International Conference on Computer Supported Collaborative Learning* (pp. 47-54). Gothenburg, Sweden: ISLS.
- Hoy, A. W. (2000, April). Changes in teacher efficacy during the early years of teaching. Paper presented at the *AERA Annual Meeting, New Orleans, LA*.
- Kurz, T. B., & Knight, S. L. (2004). An exploration of the relationship among teacher efficacy, collective teacher efficacy, and goal consensus. *Learning Environments Research*, 7(2), 111-128.
- Laferrière, T., Montane, M., Gros, B., Alvarez, I., Bernaus, M., Breuleux, A., ... & Lamon, M. (2010). Partnerships for knowledge building: An emerging model. *Canadian Journal of Learning and Technology*, 36(1).
- Leithwood (2018). *Leadership Development on a Large Scale: Lessons for Long-Term Success*. Corwin.
- Leithwood, K., & Jantzi, D. (2008). Linking leadership to student learning: The contributions of leader efficacy. *Educational Administration Quarterly*, 44(4), 496-528.
- Resendes, M., Scardamalia, M., Cordeiro, M., Massey, L., Dobbie, K., & Sirois, L. (2016). A Tri-Level Partnership to Support and Spread Knowledge Building in Ontario. In Looi, C. K., Polman, J. L., Cress, U., and Reimann, P. (Eds.). *Proceedings of the 12<sup>th</sup> International Conference of the Learning Sciences* (pp. 1249-1250). Singapore: ISLS.
- Resendes, M. & Dobbie, K. (2017). *Knowledge Building Gallery: Teaching for Deep Understanding and Community Knowledge Creation*. D. Maika, E. Hine, L. Ma, E. Heaver, & L. White (Eds.). Retrieved from <https://thelearningexchange.ca/wp-content/uploads/2017/04/Knowledge-Building-Booklet-Accessible-1.pdf>
- Ross, J. A., & Gray, P. (2006). School leadership and student achievement: The mediating effects of teacher beliefs. *Canadian Journal of Education*, 798-822.

- Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith, & C. Bereiter (Eds.), *Liberal education in a knowledge society* (pp. 67-98). Berkeley, CA: Distributed by Publishers Group West.
- Scardamalia, M., & Bereiter, C. (2003). Beyond brainstorming: Sustained creative work with ideas. *EDUCATION CANADA-TORONTO-*, 43(4), 4-8.
- Scardamalia, M., & Bereiter, C. (2014). Knowledge building and knowledge creation: Theory, pedagogy, and technology. In K. Sawyer (Ed.), *Cambridge Handbook of the Learning Sciences* (pp. 397-417). New York: Cambridge University Press.
- Schunk, D. H., & DiBenedetto, M. K. (2016). Self-efficacy theory in education. *Handbook of Motivation at School*, 2, 34-54.
- Takahashi, S. (2011). Co-constructing efficacy: A “communities of practice” perspective on teachers’ efficacy beliefs. *Teaching and Teacher Education*, 27(4), 732-741.
- Teo, C. L. (2014). Conceptual shifts within problem spaces for knowledge-building practice within a teacher community. In *Knowledge Creation in Education* (pp. 205-226). Springer, Singapore.
- Toth, P. & Ma, L. (2018). Fostering student voice and epistemic agency through knowledge building. In *Proceedings of the 21st Knowledge Building Summer Institute*; Toronto, Canada: Knowledge Building International.
- Tschannen-Moran, M., & Barr, M. (2004). Fostering student learning: The relationship of collective teacher efficacy and student achievement. *Leadership and Policy in Schools*, 3(3), 189-209.
- Tschannen-Moran, M., & Hoy, A. W. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education*, 17(7), 783-805.
- Vokatis, B., & Zhang, J. (2016). The Professional Identity of Three Innovative Teachers Engaging in Sustained Knowledge Building Using Technology. *Frontline Learning Research*, 4(1), 58-77.

## Endnotes

- (1) L’Association des directions et directions adjointes des écoles franco-ontariennes, Catholic Principals’ Council | Ontario, and Ontario Principals’ Council

## Acknowledgments

We are grateful to Knowledge Building educators and consultants, Dr. Rob Iannuzzi, Elaine Hine and Denis Maika for their time and support. Their creative ideas and insight have been invaluable to our Knowledge Building.

# Comparing the Effectiveness of Supports for Collaborative Dialogic Sense-Making with Agent-Based Models

Ilana Dubovi, Ben-Gurion University of the Negev, [dubovi@post.bgu.ac.il](mailto:dubovi@post.bgu.ac.il)  
Victor, R. Lee, Utah State University, [victor.lee@usu.edu](mailto:victor.lee@usu.edu)

**Abstract:** Collaborative dialogue has been classified by the ICAP framework (Chi & Wylie, 2014) as the highest interactive level of cognitive engagement. In this study we evaluate the balance between type of guidance and learner characteristics while collaborative learning with agent-based models takes place. Participants were students from health-care programs and were randomly assigned to one of two conditions: one group learned with agent-based models using the vicarious approach, where pairs viewed and discussed recordings of others using agent-based models; the other group explored agent-based models while discussing in pairs a set of text-based prompts. The results reveal that the vicarious-learning approach was superior to agent-based models' exploration. More detailed evaluation shows that nursing students' performance wasn't affected by the type of guidance, while less experienced in the content, non-nursing students, benefited more from vicarious guidance. Findings suggest that dialoguing while observing the dialogue of others can maximize immediate learning gains.

**Keywords:** agent-based models, vicarious learning, collaboration, ICAP framework, dialogue

## Introduction

According to Chi's ICAP framework (Interactive, Constructive, Active, and Passive), collaborative dialoguing is the highest interactive level of cognitive engagement (Chi et al., 2018; Chi & Wylie, 2014). While differentiating modes of cognitive engagement on the basis of students' overt behaviors, dialogue can be considered truly interactive only when it consists of mutual exchanges of ideas between two (or more) individuals, resulting in new ideas that neither individual knew initially nor could generate alone. Computer-based simulations and models can provide excellent opportunities to generate constructive dialogue with the aim to promote more in-depth conceptual understanding of science content, especially when complex systems are involved (D'Angelo et al., 2014). Of special interest to this study is agent-based modeling (ABM). Joint exploration of agent-based models allows students to test their ideas, negotiate, argue a position and provide justifications, compare, and revise one another's understandings based on the dynamic feedback provided by the running simulation model (Wilensky & Reisman, 2006). Yet, to achieve these potential benefits of learning with agent-based models, there is a need for learning activities that provide appropriate guidance and support interactive dialoguing (see meta-analysis by Lazonder & Harmsen, 2016).

The current study seeks to compare types of support that facilitate constructive joint dialogue by accounting for some learner characteristics within the context of an instructional unit about diabetes mellitus pharmacology for health-care professions education. We compare a vicarious learning approach of observing collaboratively a dialogue between a learner and an instructor with one involving guided exploration and explicit written prompts to have dialogue related to a set of agent-based models. This study is intended to question the efficacy of vicarious learning with agent-based models by comparing it with guided exploration.

## Vicarious learning

Learning from observing, or learning vicariously, was first proposed by social psychology research (Bandura, 1969) to evaluate learning by imitating and modeling someone's else behaviors and actions. Bruner (1986) stated that "most of our encounters with the world, are not direct encounters" (p. 122), which would seem to imply that it is possible to learn through mechanisms other than primary or firsthand experience. Such social learning is effective without the need for the observer to experience feedback directly. Bandura's introduction of the idea – vicarious reinforcement – was based on classic studies in which children were seen to imitate the aggressive behavior modeled by adults as they assaulted Bobo dolls (Bandura, Ross, & Ross, 1963).

Vicarious learning has been also explored in area of learning of cognitive behaviors as asking questions (Rummel & Spada, 2005), doctor-patient communications skills (Stegmann, Pilz, Siebeck, & Fischer, 2012), and acquiring complex cognitive skills (Chi, Roy, & Hausmann, 2008). Initial research on the differences between direct participation in a task and simple observation of it being performed was conducted by Schober and Clark (1989), in which pairs of participants communicated about a task that involved sequencing figures correctly. Simply overhearing the dialogue, rather than participating directly, resulted in poorer performance, leading

Schober and Clark to conclude that the direct participants had an advantage by being able jointly to construct a common ground of comprehension. However, in a replication of Schober and Clark's study (Chi et al., 2008), in which the vicarious learners were required to be active, by self-explaining aloud while overhearing a tutor-tutee interaction, vicarious learners did just as well as tutees who interacted with a tutor.

To further explore vicarious learning, Chi and her colleagues tested different conditions involving different vicarious aspects – observing monologue videos alone / observing dialogue videos alone / observing monologue videos collaboratively / observing dialogue videos collaboratively (Chi et al., 2008). The results suggest that vicarious learning through observation doesn't have to be passive but might instead be considered interactive when the conditions of *collaboratively* overhearing and observing *dialogue* between a learner and an instructor are taking place. Dialogue that incorporates mutual exchanges of ideas between individuals will result in new ideas that neither individual knew initially nor could generate alone. Therefore, the ICAP theoretical framework, which categorizes different modes of active learning, defines joint dialoguing as the highest interactive level of cognitive engagement (Chi & Wylie, 2014). The advantage of designing vicarious learning with the dual discourse approach, collaboratively observing a dialogue, is twofold. First, observing collaboratively facilitates interactions between peers that offer an opportunity for exchanges of ideas and argumentation processes which foster learning (Schwarz, Neuman, & Biezuner, 2000). Second, observation of dialogues over monologues makes it possible for vicarious learners to overhear the tutee's questions and struggles and to reflect on their own mental models using the tutor's feedback as well. As a result, learning gains of collaborative observations of dialogues were similar to learning gains following face-to-face human tutoring (Muller, Sharma, & Reimann, 2008). This study is aimed at adding to the existing domain of knowledge on the effectiveness of vicarious learning (Chi et al., 2018) by comparing pairs collaboratively observing videos vicariously with pairs following guided exploration coupled with explicit prompts to dialogue about agent-based models within the antidiabetic pharmacology education context.

### Agent-Based Modeling environment

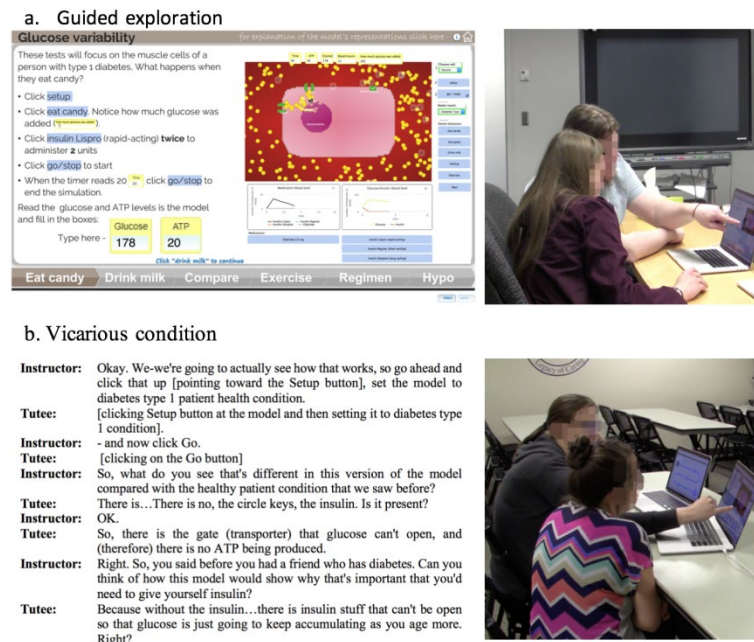
ABM is a computational modeling paradigm that emphasizes multi-level examination of complex, multi-agent systems. The ABM paradigm encodes the behavior of individual agents in a small set of simple rules so that we can specify and observe the results of these agents' individual actions and interactions. Learning through this approach focuses on entities and their actions (also called the micro level of the system), such as movement, interactions, and global flows (also called the macro level of the system), and allows students to comprehend parallel processes by which emergent phenomena form (Wilensky & Resnick, 1999). Exploration of agent-based models encourages causal emergent thinking in connecting individual behaviors with systemic patterns, thus helping students learn various scientific concepts effectively (i.e., Samon & Levy, 2017).

Yet, to achieve these potential benefits of agent-based models exploration, research has provided conclusive evidence that students need support and guidance (D'Angelo et al., 2014; Kirschner et al., 2006). However, the educational literature presents varied types and levels of guidance, from prompts and cues to specific and direct explanations of how to perform an action (Lazonder & Harmsen, 2016). There is an ongoing debate, the "assistance dilemma", about what type of guidance is adequate and whom it fits (Koedinger & Alevan, 2007, p. 261). On one side of this argument are those advocating direct instructional guidance as beneficial for novice learners (e.g., Kirschner et al., 2006); on the other side are those suggesting that less heavily guided instructions may be more productive, encouraging concept learning and transfer (e.g., Koedinger & Roll, 2012). Hmelo-Silver and her colleagues challenged this notion by suggesting that instead of contrasting these dichotomous positions on the continuum of guidance level, the questions to be asked are "under what circumstances do these guided inquiry approaches work, what are the kinds of outcomes for which they are effective . . . and what kinds of support and scaffolding are needed for different populations and learning goals?" (Hmelo-Silver, Duncan, & Chinn, 2007, p. 105).

In fact, regarding the exploration of agent-based models, it is still unclear how to balance the instructional guidance. On the one hand, agent-based models are exploratory models that are designed so that students can discover things by experimenting within a microworld; on the other hand, without providing explanations, understanding of complex systems won't become explicit but might remain implicit. Vitale, McBride, and Linn (2016) compared two forms of automated guidance – direct explanations and prompts to motivate learners (knowledge integration guidance) – to support learning of complex systems with agent-based models. They found that directed guidance for agent-based models exploration produced higher immediate learning gains than minimal guidance did. However, recent studies by Jacobson et al. (2017) and Levy et al. (2018, p. 2) on agent-based models' exploration for complex systems learning suggested that minimal guidance, or a "constrained discovery", can be valuable to conceptual learning.



The current study focuses on science learning, particularly pharmacology learning, among health-care professionals in a higher education setting. Moreover, we aimed to support science learning within health-related education, which has so far occupied only a modest space in the learning sciences. Therefore, we present the Deep Dive into Diabetes (DDD) agent-based learning environment, which was constructed with the NetLogo modeling platform (Wilensky, 1999). The proposed agent-based models, which simulate biochemical processes that represent the relevant anatomy of glucose equilibrium and the mechanisms of drug action, were constructed in a previous study (Dubovi, Dagan, Mazbar, Nassar & Levy, 2018). To explore the effect of guidance level, the DDD learning environment was constructed with two different levels of learning activities that support learning with agent-based models (Figure 1): (1) collaborative observation of a video tutorial, a vicarious approach; and (2) collaborative exploration and experimentation with agent-based models, a guided-exploration approach.



**Figure 1.** DDD learning environment: (a) Guided-exploration condition: screenshot, and photo of students learning with the environment; (b) Vicarious condition: example of conversation that students were overhearing while learning with the video tutorial, and photo of students learning with the environment.

## Methods

### Research design

This study employed a quasi-experimental pre- and post-test design with two comparison conditions to explore the effect of guidance level on inquiry learning with agent-based models. We conducted a quantitative analysis of questionnaires of all participants (70) and video analysis of 11 randomly selected dyads of students (22).

### Participants and procedure

The participants were 70 health-care undergraduate students (40 students from a nursing program, 23 from a nutrition program, and 7 from a health education program) who volunteered to participate in this study. Most participants were females (61), and the average age was  $23 \pm 4.85$  years. Participants were randomly assigned to participate in pairs in one of two conditions: a guided-exploration (34) or a vicarious condition (36). Both conditions involved active and continuous dialogue in pairs as participants tried to interpret what was happening in the DDD learning environment. In the first condition, students were required to learn about diabetes through guided exploration of agent-based models (the guided-exploration condition); in the second condition, students learned about diabetes by observing instructor-learner dialogue and by exploring agent-based models (the vicarious condition).

There were no statistically significant differences in demographic characteristics such as grade point average (GPA) ( $\chi^2=2.58, p=0.46$ ) or prior knowledge about diabetes ( $t=1.36, p=0.17$ ) between the two conditions.

On average, students spent about 50 minutes learning with the vicarious-instruction and 52 minutes learning with the guided-exploration-instruction method.

## Instruction design and materials

While learning with the DDD environment, students in the both groups were asked to use information from an agent-based model to solve problems about diabetes and antidiabetic pharmacology content. The *vicarious condition* group received guidance by observing a video tutorial of an instructor and a learner while exploring the agent-based models. A staff member at a university who was an advanced novice on diabetes and who was familiar with the NetLogo modeling system served as the instructor, and a graduate student who had passing familiarity with diabetes and the modeling system was the learner. This was an unscripted conversation. The study participants were able to overhear the dialogue between learner and instructor that included the instructor's explanations of agent-based models' representations; what to pay attention to while the model is running and how to interpret the model's output; and questions raised by the student which were followed by the instructor's feedback (Figure 1b). Similar to Chi et al.'s study, students could pause, reverse, and skip portions of the video; to enhance participants' active cognitive engagement while observing, we also followed Chi and her colleagues' vicarious learning design by asking learners to solve together several sub-problems which served as landmarks to the video and prompted interactive dialogue between the participants (Chi et al., 2008). The *guided exploration condition* group received less guidance while directly exploring the agent-based models. The scaffolds included information on how to set up the agent-models and on which graphs and monitors to pay attention to. As in the vicarious condition, to facilitate dialogue between pairs, students were asked to talk with one another and to solve sub-problems while exploring the agent-based models (Figure 1a). Their solutions to the problems were entered into a single shared computer. We used learning activities that we developed in a previous study (Dubovi et al., 2018).

## Data collection instruments

### Diabetes knowledge questionnaire

The Diabetes knowledge questionnaire was adapted from the Pharmacology Diabetes Mellitus questionnaire (PDM) developed in a previous study (Dubovi et al., 2018). The questionnaire consists of nine questions (7 multiple-choice, 2 open-ended), and evaluates understanding of biochemical glucose equilibrium, glucose disequilibrium (i.e., diabetes type 1 and diabetes type 2), and medications actions. Analysis of the Diabetes knowledge questionnaire using Cronbach's alpha yielded an internal consistency score of 0.68, which was similar to our previous report ( $\alpha = 0.71$ ) and can be considered acceptable.

Responses to the questionnaire were coded as either correct or incorrect, and the total score was calculated as the percentage of correct answers. In addition, the items on the Diabetes knowledge questionnaire were scaled by level of difficulty: four items were coded as the most difficult, and five items as the least difficult. Level of difficulty for each item was determined based on the percentage of students who correctly answered it on the post-test. Although students completed the pre- and post-tests questionnaires as individuals, their learning process was nested within their collaboration as a pair. Prior knowledge varied greatly between students as individuals as well as within pairs. To account for these variations, the analysis invoked multi-level modeling. Owing to the pre-test/post-test design used in this study, our data analysis encompassed repeated measures on individuals over time. Consequently, a three-level structure arose: both test times (Level 1) were clustered within students (Level 2), which were nested within dyads (Level 3). Although multi-level models quantified the variance across pairs, the focus of the study was on at the individual student level.

### Video recordings

To assess the learning process, we recorded students' discourse and interactions with the agent-based models using screen-recording software and a separate standing video camera. For the analysis, we randomly chose 11 pairs (22 individual students); five pairs learned with the guided-exploration condition, and six pairs learned with the vicarious condition. We evaluated the frequency of accuracy of students' ideas and explanations as they learned with the DDD environment. To generate students' ideas accuracy, discourses and dialogues were carefully transcribed, iteratively reviewed, and coded in terms of the ideas, explanations, and statements expressed. This approach to the selection of knowledge elements and ideas is comparable to the approach documented in Sherin, Krakowski, and Lee (2012) and Minstrell (1982). From this, transcript excerpts were identified to illustrate some of the dialogues and ideas expressed.

## Results

### Diabetes knowledge questionnaire scores

Multilevel model analysis was conducted to determine the effect of pre-test scores, experimental condition (vicarious vs. guided exploration), and the student's program of study (nursing vs. other health-care field) on post-test outcomes, both independently and through examining the interactions between them. Moreover, multilevel model analysis also considers the random effect of individual and pair characteristics on factors' interactions. As shown in Table 1, after adjusting for differences between individuals and dyads, the overall post-test score for the Diabetes knowledge questionnaire was the sum of the intercept (37.318). A significant interaction of Time × Experimental condition indicates that the vicarious-condition pre-test to post-test learning gains were significantly higher than those for the guided condition (34.22 to 67.47 vs. 40.23 to 61.84).

Examining the level of the Diabetes knowledge questionnaire items independently, being in a nursing program versus another health-care program moderates the effect of experimental condition within the most difficult items (Table 1). More specifically, students from non-nursing programs made significantly higher learning gains when participating in the vicarious condition compared with the guided-exploration condition (8.93 to 60.71 vs. 18.53 to 38.84, respectively; Figure 2), whereas students from the nursing program gained knowledge from learning with both conditions, with no significant differences (Figure 2). This effect was true only for the most difficult questionnaire items; for the least difficult items, the interaction Time × Experimental condition was insignificant, meaning that students showed similar learning gains from learning with the vicarious (nursing: 51.60 to 65.39; non-nursing 57.84 to 71.63) and the guided-exploration conditions (nursing: 52.13 to 65.91; non-nursing 58.37 to 72.16).

**Table 1: Three-level nested random-intercepts multilevel model predicting students' Diabetes knowledge questionnaire post-test scores**

|  | <b>Overall Questionnaire</b> | <b>The Most Difficult Items</b> | <b>The Least Difficult Items</b> |
|--|------------------------------|---------------------------------|----------------------------------|
| <b>Fixed Effects</b>                         | <b>Estimate (SE)</b>         | <b>Estimate (SE)</b>            | <b>Estimate (SE)</b>             |
| Intercept                                    | 35.279 (3.839) ***           | 18.535 (5.764)**                | 51.762 (4.389)***                |
| Time: Post vs. Pre                           | 21.609 (3.187) ***           | 20.313 (6.346)**                | 13.786 (2.771)*                  |
| Experimental condition: Vicarious vs. Guided | -6.010 (4.543)               | -9.602 (8.428)                  | -0.941 (4.669)                   |
| Program: Nursing vs. Other                   | 8.671 (3.859)*               | -0.288 (7.817)                  | 7.264 (4.706)                    |
| <b>Interactions:</b>                         |                              |                                 |                                  |
| Time × Experimental condition                | 11.647 (4.444) **            | 31.473 (9.290)***               | -                                |
| Time × Program                               | -                            | 28.298 (8.722)**                | -                                |
| Experimental condition × Program             | -                            | 2.716 (11.029)                  | -                                |
| Time × Experimental condition × Program      | -                            | -24.403 (12.304)*               | -                                |
| <b>Random Effects</b>                        | <b>Var</b>                   | <b>Var</b>                      | <b>Var</b>                       |
| Dyads  | 79.364                       | 65.575                          | 231.849                          |
| Participants within Dyad                     | 53.560                       | 79.733                          | 6.196                            |
| Residual                                     | 172.678                      | 322.206                         | 268.727                          |

Note: Sample size is 35 dyads made up of 70 participants with 140 total observations.

Each model contains only significant interactions.

\*p<0.05, \*\*p<0.01, \*\*\*p<0.001.

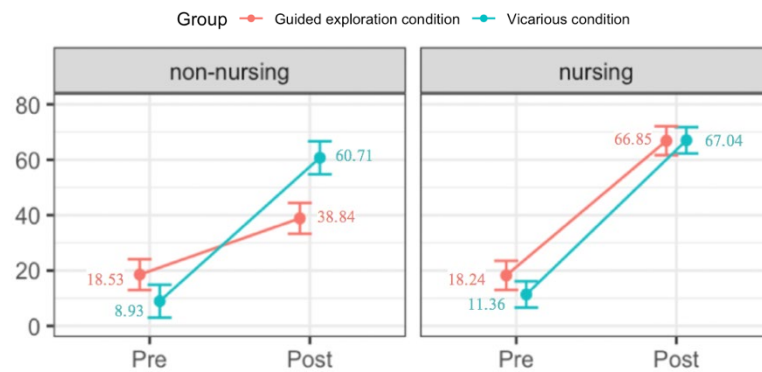


Figure 2. Visualization of significant interaction effect of Diabetes knowledge questionnaire’s most difficult items, by experimental condition, among non-nursing students; and the non-significant interaction effect of the most difficult items by experimental condition, among nursing students. Depicted are estimated marginal means  $\pm 1$  standard error of the mean (SEM).

## Learning process

To further examine the learning process that characterized learning with the DDD environment, we analyzed the video-recording data of 11 pairs to evaluate students’ number and quality of different ideas exchanged between the partners. About 230 ideas and explanations were identified for five pairs who learned with the guided-exploration condition (average number of ideas for each pair was 48) and 305 for six pairs who learned with the vicarious condition (average number of ideas for each pair was 51). The following short discourse of a student pair assigned to the guided-exploration condition illustrates the expression of some accurate and less accurate ideas (All names are pseudonyms to protect participants’ identities.):

- Jessica: Go. Now I click Go button. [setting the model to the diabetes type 1 condition]
- Michelle: So, we should have him eat candy.
- Jessica: Eat candy [clicking on the button which says “eat candy”]. It will be so much glucose.
- Michelle: So [observing the agent-based model], nothing is getting through [the muscle cell] at all.
- Jessica: Yeah, but, because there is no insulin.
- Michelle: Ahhh . . . there’s no Insulin. So, like, if they drink milk or eat pasta, or fasting, it won’t matter, doesn’t it?
- Jessica: There’s nothing. So, you do not need to exercise. There’s no insulin. They still need insulin to be alive.
- Michelle: Yeah, [without insulin] they are unable to make ATP.
- Jessica: Mm-hmm. So, there’s no energy and then what?!

During this episode, Jessica and Michelle articulate three basic ideas that are critical to understanding the pathological processes related to glucose equilibrium as part of the action of antidiabetic drugs. The first is insulin’s role in regulating glucose metabolism: if there is no insulin, then “nothing is getting in”, meaning that without insulin, glucose can’t enter the GLUT4 transporters on muscle cells’ membranes. This idea is correct for the muscle and lipid tissues, where glucose metabolism is mediated by insulin hormones. The second idea implicates the pathological definition of type 1 diabetes: “there is no insulin”. Here Jessica expresses a nominal fact: that in type 1 diabetes there is no insulin. Next, the relationship between glucose molecules and ATP production is noted, namely the cellular respiration metabolic pathway (Glycolysis). Jessica summarizes here her third idea, that when there is no glucose within the cells, the ATP production process is impaired, which means that “there’s no energy” production. This explanation is only partially correct because fatty acids and amino acids can be used for ATP production.

A close look at the correctness of students' ideas revealed that students who learned with the guided-exploration condition (five pairs) expressed 41 inaccurate ideas, whereas students who learned with the vicarious condition (six pairs) expressed only seven inaccurate explanations. Most of the inaccurate explanations with the guided-exploration condition resulted from misinterpreting what to pay attention to while exploring the agent-based models.

## Discussion

The most interesting result is the advantage of vicarious learning for immediate learning achievements over the guided-exploration condition. One possible explanation for this effect is that learning with the guided-exploration condition triggered more inaccurate ideas and explanations, even though the number of ideas that students expressed was not affected by the experimental condition. Interestingly, a more precise analysis reveals that there are differential instructional effects across students' characteristics. Evaluation of the more difficult items of the Diabetes knowledge questionnaire shows that nursing students' performance wasn't affected by the type of guidance but instead benefited equally from both the vicarious-learning guidance and the guided-exploration support. However, students from the non-nursing programs, namely nutrition and health-education programs, made higher learning gains within the more difficult items of the Diabetic knowledge questionnaire following learning with the vicarious approach. Hence, non-nursing students benefited more from the vicarious approach than from the guided-exploration one. Although our results show no significant difference in prior diabetes knowledge between the nursing and non-nursing students, a plausible explanation for the difference is related to nursing students' prior experiences with the pharmacology topics from their clinical practices and curriculum. Whereas students from nutrition and health-education programs are focused on food science and on health promotion and less on pharmacology concerns, nursing training notably emphasizes nurses' need for accountable and responsible medication management, which is crucial for nursing practice (Khan & Hood, 2018). We propose that this acknowledgment of pharmacology importance as part of a professional identity better prepared nursing students for learning and for making meaning from any support that was available to them, whether vicarious or guided exploration.

As stated in the introduction, the ICAP framework predicts that the more active students are in their learning activities, the better their learning outcomes will be. According to this framework, both conditions of the current study evoked active engagement using both constructive activities (e.g., self-explanations, predictions, and model exploration) and interactive mode involving joint dialogue (Chi & Wylie, 2014). The main practical implication of this study is that learning with tutorials can be scaled up and maximized when visual displays encourage dialogue between instructor and students and when learners solve problems while observing and overhearing this dialogue collaboratively. As we showed, by evoking an active level of engagement using dialoguing, vicarious instructional design to support learning with agent-based models can be at least as efficient as the guided exploration of agent-based models. Vicarious learning provides a unique modeling opportunity for students to learn how to explore agent-based models and enables them to reflect on their own process of learning. Our results show that less-experienced students can especially benefit from the vicarious approach to support learning with agent-based models until they achieve a certain level of expertise.

This study has several limitations, specifically, the current study evaluated only immediate learning effects followed by one short intervention, using a single population. Therefore, the advantage of the vicarious approach over the guided exploration of modeling systems should be further explored. Open exploration of agent-based models is important for concept construction; our findings underline what makes working with agent-based models more or less challenging with different instruction types for different populations.

## References

- Bandura, A. (1969). *Principles of behavior modification*. New York, NY: Holt, Rinehart & Winston.
- Bandura, A., Ross, D., & Ross, S. (1963) Vicarious reinforcement and imitative learning. *Journal of Abnormal and Social Psychology*. 67(6), 601–607.
- Bruner, J. S. (1986). *Actual minds, possible worlds*. Cambridge, MA: Harvard University Press.
- Chi, M. T., Adams, J., Bogusch, E. B., Bruchok, C., Kang, S., Lancaster, M., . . . & Wylie, R. (2018). Translating the ICAP theory of cognitive engagement into practice. *Cognitive Science*, doi:10.1111/cogs.12626.
- Chi, M. T., Roy, M., & Hausmann, R. G. (2008). Observing tutorial dialogues collaboratively: Insights about human tutoring effectiveness from vicarious learning. *Cognitive science*, 32(2), 301–341.
- Chi, M. T., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, 49(4), 219–243.
- Chi, M. T., Roy, M., & Hausmann, R. G. (2008). Observing tutorial dialogues collaboratively: Insights about human tutoring effectiveness from vicarious learning. *Cognitive Science*, 32(2), 301–341.

- D'Angelo, C., Rutstein, D., Harris, C., Bernard, R., Borokhovski, E., & Haertel, G. (2014). *Simulations for STEM learning: Systematic review and meta-analysis*. Menlo Park: SRI International.
- Dubovi, I., Dagan, E., Mazbar, O. S., Nassar, L., & Levy, S. T. (2018). Nursing students learning the pharmacology of diabetes mellitus with complexity-based computerized models: A quasi-experimental study. *Nurse education today*, *61*, 175-181.
- Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: a response to Kirschner, Sweller, and Clark. *Educational Psychologist*, *42*(2), 99–107.
- Khan, E. U., & Hood, P. A. (2018). Nurses' perspectives on pharmacology: why, what and at which point of the curricula should education be delivered? *British Journal of Nursing*, *27*(10), 546–553.
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*(2), 75–86.
- Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review*, *19*(3), 239–264.
- Koedinger, K. R., & Roll, I. (2012). Learning to think: Cognitive mechanisms of knowledge transfer. In K. J. Holyoak & R. G. Morrison (Eds.), *The Oxford handbook of thinking and reasoning* (pp. 789–806). New York, NY: Oxford University Press.
- Jacobson, M. J., Markauskaite, L., Portolese, A., Kapur, M., Lai, P. K., Roberts, G., (2017). Designs for learning about climate change as a complex system. *Learning and Instruction*. *52*, 1–14.
- Lazonder, A. W., & Harmsen, R. (2016). Meta-analysis of inquiry-based learning: Effects of guidance. *Review of Educational Research*, *86*(3), 681–718.
- Levy, S. T., Peleg, R., Ofek, E., Tabor, N., Dubovi, I., Bluestein, S., & Ben-Zur, H. (2018). Designing for discovery learning of complexity principles of congestion by driving together in the TrafficJams simulation. *Instructional Science*, *46*(1), 105–132.
- Minstrell, J. (1982). Facets of students' knowledge and relevant instruction. In R. Duit, F. M. Goldberg, & H. Niedderer (Eds.), *Proceedings of Research in physics learning: Theoretical issues and empirical studies* (pp. 110–128). Keil, Germany: The Institute for Science Education (IPN).
- Muller, D. A., Sharma, M. D., & Reimann, P. (2008). Raising cognitive load with linear multimedia to promote conceptual change. *Science Education*, *92*(2), 278–296.
- Rummel, N., & Spada, H. (2005). Learning to collaborate: An instructional approach to promoting collaborative problem solving in computer-mediated settings. *Journal of the Learning Sciences*, *14*, 201–241.
- Samon, S., & Levy, S. T. (2017). Micro–macro compatibility: When does a complex systems approach strongly benefit science learning? *Science Education*, *101*(6), 985–1014.
- Sherin, B., Krakowski, M., & Lee, V. R. (2012). Some assembly required: How scientific explanations are constructed in clinical interviews. *Journal of Research in Science Teaching*, *49*(2), 166–198. doi:10.1002/tea.20455
- Schwarz, B. B., Neuman, Y., & Biezuner, S. (2000). Two wrongs may make a right . . . if they argue together! *Cognition and Instruction*, *18*(4), 461–494.
- Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and observers. *Cognitive Psychology*, *21*, 211–232.
- Stegmann, K., Pilz, F., Siebeck, M., & Fischer, F. (2012). Vicarious learning during simulations: Is it more effective than hands-on training? *Medical Education*, *46*(10), 1001–1008.
- Vitale, J. M., McBride, E., & Linn, M. C. (2016). Distinguishing complex ideas about climate change: Knowledge integration vs. specific guidance. *International Journal of Science Education*, *38*(9), 1548–1569.
- Wilensky, U. (1999). NetLogo. Center for Connected Learning and Computer-Based Modeling. Northwestern University, Evanston, IL.
- Wilensky, U., & Reisman, K. (2006). Thinking like a wolf, a sheep, or a firefly: Learning biology through constructing and testing computational theories—an embodied modeling approach. *Cognition and Instruction*, *24*(2), 171–209.
- Wilensky, U., & Resnick, M. (1999). Thinking in levels: A dynamic systems approach to making sense of the world. *Journal of Science Education and Technology*, *8*(1), 3–19.

## Acknowledgments

This work was supported by AIS - Teaching with Technology Innovation Grant, Utah State University. We are especially grateful to Dr. Sarah Schwartz for her assistance with the statistical analysis.

# Using Voice Assistant Skills in Family Life

Diana Beirl, UCL, UK, dianabeirl01@gmail.com

Yvonne Rogers, UCL, UK, y.rogers@ucl.ac.uk

Nicola Yuill, University of Sussex, nicolay@sussex.ac.uk

**Abstract:** Voice assistants are increasingly prevalent in family life, being used, for example, for listening to music, finding out information, asking jokes and playing games together. However, little research shows how such technology influences dynamic family interactions in the home over time. An in-the-wild study was conducted in six family homes over three weeks. An Alexa, with a number of skills, was set up in each home for the families to use. The findings showed differences in use over time. To begin, family cohesion behavior and family rituals were most prevalent. At the end of the study, the skills were found to motivate distinct family interaction patterns: including more collaboration to manage Alexa and scaffolding of children's interactions with Alexa, given developmental differences in users' grasp of Alexa's capabilities and limitations. We discuss how voice assistants support different interaction patterns and potentially, offer different learning opportunities.

## Introduction

A new technology in many family homes is a voice assistant, such as Amazon's Alexa or Google Home. These devices are often placed in the kitchen or living room, enabling all to interact with it; providing an additional source of entertainment for the family, for example, telling jokes, playing music, keeping track of time and playing games. Recently, Amazon's Alexa has expanded its repertoire by enabling users to download a diversity of 'skills'. These are voice-driven capabilities that are intended to provide a more personalized experience. For example, "Open the Magic Door" is an adventure skill that allows users to choose their path in a story by selecting different options through the narrative. Another one, "Kids court", allows families to settle arguments in an Alexa-run court while learning about the law.

Many of the skills are designed to support multiple users taking part at the same time, offering the potential for families to play together. This kind of social interaction is further encouraged by the affordance of the physical device – that differs from other virtual voice assistants, found on phones and laptops, insofar as their physical presence affords joint ownership and use – similar to other domestic devices, such as the radio, TV and toaster. Little is known, however, about how this combination of device and interactive user experiences are used by families in their own homes.

Playing games and music together can contribute to family bonding and building social and emotional connections (Boer *et al.*, 2014). However, the pervasive uptake of technology in the home is often seen as disrupting such shared engagement, arguably resulting in less face-to-face conversation (Turkle, 2015). But might the introduction of virtual assistants, as shared devices, provide both new opportunities and challenges for the ways families interact in the home? In particular, could the new types of skills, that run on the voice assistants, support these kinds of family interactions, by enabling new repertoires and learning to emerge? If so, how do they manifest themselves? Conversely, could interacting with them lead some family members to dominate while others get frustrated? Could they even promote sibling rivalry and family arguments?

We began to address these questions by conducting an in-the-wild study observing how the voice assistant, Alexa, was used and accommodated in family homes over a period of three weeks. Three different types of skills were downloaded onto it, for the categories of music, storytelling and games. These skills differ in the role that Alexa plays. In providing music, Alexa acts purely to provide content for the user to enjoy either for themselves or as a shared experience. The user usually drives what is played, either independently or in consultation with other people present. In storytelling, Alexa has a greater role in providing potentially immersive content. An audio story is presented with options that lead to different outcomes, hence providing more interaction between Alexa and the users. Games, such as Kids Hub, support a diversity of interaction styles, that have more potential for multiple users interacting with Alexa, as well as parent-child pairs.

The aim of our study was to explore how families learn the new skills and appropriate them into their life. Video observations and interviews were conducted to collect data for how and how often families used the skills. The findings from the study were analysed in terms of different kinds of family interactions exhibited across families. A number of different types of interactions were identified that facilitated family cohesion, bonding and empathy. Some entailed helping children to learn how to play, ask questions and take turns. Others encouraged rituals, routines and rivalry often exhibited in family life. We discuss how families adapt and exploit

the constraints imposed by voice assistants, like Alexa, to play, talk and learn and conversely, how different Alexa skills shape family interactions.

## Background

The home provides an informal learning environment where it is commonplace for parents to buy technological devices, video games and electronic toys for their children to learn with. Family practices, such as how technology is used for play in the home, have an influence in developing children's competencies with technology (Plowman *et al.*, 2010). Children's experiences with technology are influenced by specific family contexts and family culture (Weisner, 2002). But how effective is home technology in supporting learning? Plowman *et al.* (2012) conducted a two-year long observation of families, with pre-school children, on use of technology in the home. They identified four main areas of learning which can be supported by technology: (i) acquiring operational skills such as learning how to click buttons and learning about cause and effect situations, (ii) expanding general knowledge such as early literacy and numeracy, (iii) increasing self-efficacy for curiosity-driven learning and (iv) learning about technology in everyday life.

Based on an ethnographic study of family routines, Davidoff *et al.*, (2005) forewarned how smart home devices may negatively impact the balance of control in the management of family life, such as leisure planning. However, studies of the introduction of novel technology into family homes has been found to impact family behavior in a variety of ways. For example, Ganesh *et al.* (2014) found that the deployment of augmented reality technology in a family home setting was able to positively distract pre-school age fussy eaters from their dislike of green vegetables but also negatively distracted their siblings. The technology worked by removing the fussy eater's focal attention from the peas on their plate and instead switching their attention to observing and playing with animated lights projected onto it. Instead of simply refusing to eat the peas – as they normally did – the augmentation encouraged the young children to view the plate of food in a new way, and in doing so, shift their focus from eating as a primary activity. However, when parents and siblings also took part in encouraging the fussy child to eat their peas, the new technology sometimes triggered a negative response, causing sibling jealousy and competition. Hence, the context had a positive effect on the fussy eater but also impacted on the social dynamics of the family.

It has also been found that technology designed for the home can engage all family members. Plowman *et al.* (2010), for example, observed how parents and older siblings contribute to young children's engagement with home technology through guiding and scaffolding their interaction through spoken language, gestures and expressions. Conversely, school-age children can take more leading roles: a study of parent-child shared reading on tablets versus paper books reported that school-age children sometimes dominated when reading from screens compared to paper (Yuill & Martin, 2016). Furthermore, device type affected the nature of family interaction: using a screen altered the pair's physical positioning and the interaction quality was less warm than for paper books. Parents may also introduce a narrative through the technology that may be of particular motivational interest to their child. For example, Bhömer *et al.* (2010) designed a photo sharing device intended as a table centerpiece where personal photos downloaded from Facebook could be shared with the family at mealtimes. They found how different family members used the device as an opportunistic display of affection to each other, saying how much they loved particular photographs that popped up. They also found it was used to highlight and reinforce family ties, through the types of conversation that were triggered.

Short *et al.* (2017) found greater intergenerational equity of participation with a robot assistant for activities with distinct participant roles. Reeves and Nass (2006) have conducted much research investigating how people approach and treat computers. They will often confide in them more and anthropomorphize them, treating them as if they had human-like qualities. Other research has shown how people respond to technologies as though they were human even if they know they are not (Fong *et al.*, 2003). Additionally, people have been shown to attribute personalities to technology and to apply similar politeness norms from human conversation to such interactions (Nass *et al.*, 1999).

Virtual Assistants add a new dimension to how home technology can play out in family life. Not only can they take over the control of certain everyday activities, such as reminding users about events, they can also act as ambient conversational 'partners'. The nature and equity of family interactions is also likely to be influenced by the structure of different activities provided by virtual assistants. Young children have judged various virtual assistants to be friendly and trustworthy, and to vary in intelligence (Douceff, 2017). Druga *et al.*, (2017) found that younger children (3-4 years old) experienced difficulty interacting with conversational and chat agents, resulting in them becoming restless. Sometimes, they would try to reword their questions or speak slower. While people so far have used virtual assistants primarily to seek information, they have also been shown to contribute to humor in social situations and interruptions in human conversation (Porcheron *et al.*, 2018). Multi-member households and particularly those with children are more likely to personify voice



assistant devices. However, a study investigating virtual assistants in multi-party conversation demonstrated a number of problematic features which affect the flow of normal social interaction (Purinton *et al.*, 2017). This included the need to repeat and refine queries which were not understood by the virtual assistants, and enforced silence by conversation participants so that the virtual assistants could better understand their queries. However, these devices also enabled collaboration, in that any member could interact with the device at any given moment. Family members also tended to reason and reflect about the queries that were made. Porcheron *et al.* (2018) also examined how voice assistants were used by families in their own homes. They conducted a conversational analysis of some audio recordings that revealed subtle cues and mechanisms used during conversations. They report how a family's interaction with the Amazon Echo is seamlessly interwoven with other ongoing activities, for example, at family mealtimes where parents are at the same time trying to get their child to eat their food. They point out how our conversations with each other and voice-assisted technologies interleave in nuanced ways, rather than being separate conversations between the family or the family and the device, that jump from one to another.

A recent panel discussion on voice assistants (Kaye *et al.*, 2018) identified as important research foci the ways that families interact with voice assistants, and how users understand their capabilities and limitations. Some answers are provided by Sciuto *et al.* (2018) who analyzed Alexa voice logs for 75 users over a period of about a year, and also interviewed a number of Alexa-owning families with young children under 4. However, to understand the roles that Alexa plays in family interaction, we also need to observe audiovisual dynamic interactions in families with varying compositions, and children across a wider age range, preferably over a period of time and in relation to a range of different voice assistant skills that might engage users differently. The aim of our research was to explore how families approach and interact with different skills downloaded onto Alexa in their own home. Our focus was on how parents and children engage with different skills when they are together, across a range of ages and family types, including only children and those with siblings.

## Method

An in-the-wild study was conducted in six family's homes. The families were given an Alexa to use at home over a period of 3 weeks. They were visited at the beginning and end of the study period when all or most of the family could be present. Ethical approval was obtained from the University Ethics Committee and informed written consent for self and children was gained from parents, as well as assent from the children. They were told that their conversations with Alexa would be recorded for subsequent analysis. All families were given the opportunity to review the Alexa recordings before analysis in case they wanted to remove any items: none chose to remove items.

## Participants

Participants were recruited through flyers, social media, word of mouth and snowballing in south-east UK. The details of the families who participated in the study are shown in Table 1. All had children in the age group of 2-13 years. A broad age range of children was selected to increase the potential for a wider set of interactions to be revealed. Five of the families had used a voice assistant at home and the other was familiar with them.

## Skills

Three skills were pre-loaded onto the voice assistants. These were (i) Music, (ii) Pac-Man stories and (iii) Kids Hub. The families with previous experience were familiar with playing music but did not have experience with the other two skills. The skills were all highly-rated on Amazon and appropriate for the wide age range of participating children. They were chosen for the types of interaction they offered and the degree to which multiple family members would likely participate in using them. (i) Music is essentially providing content, as easily for an individual as for a group. Individual use is supported because music interest likely varies across our wide age range. However, music could also potentially encourage playful interaction in the form of dancing and singing, given the shared audio experience for anyone in the room at the time. (ii) Pac-Man stories is a spoken interactive adventure style game. It similarly provides content accessible for an individual, but has more potential for interaction, because Alexa invites the player to choose alternate paths. The game incorporates sound effects and music, thus supporting the user to become immersed in the story. As with music, this may also have the potential to encourage participation in the form of dancing or engagement from family members not primarily using the skill. (iii) Kids Hub supports a diverse range of more interactive participation, allowing users to choose from a variety of activities including songs with actions, Tongue Twisters, Pictionary and Kiddy Olympics. These activities can be readily used by more than one person and are appropriate for families with single and multiple children.

Table 1: Details of families who participated in the study

|                 | Children        | Age      | Parents  | Previous voice assistant |
|-----------------|-----------------|----------|----------|--------------------------|
| <b>Family 1</b> | Boy             | 6        | Mum, dad | Yes – Google Home        |
| <b>Family 2</b> | Boy             | 11       | Mum, dad | Yes – Amazon Alexa       |
| <b>Family 3</b> | Boy             | 13       | Mum      | Yes – Amazon Alexa       |
| <b>Family 4</b> | Boy, girl, girl | 6, 9, 11 | Mum, dad | Yes – Google Home        |
| <b>Family 5</b> | Boy, boy        | 7, 10    | Mum, mum | No                       |
| <b>Family 6</b> | Girl            | 2        | Mum      | Yes – Amazon Alexa       |

## Procedure

The researcher visited each family home for about an hour and set up the Amazon device in a place chosen by the family – the living room or the kitchen. She explained how to use the three skills. The families were free to explore them. A camera was set up in the corner of the room to record the interactions during her visit. Afterwards, the family members were asked about their reflections of Alexa and how they imagined it would be used in the home. The families were asked to explore the skills for a week. They were then sent a link to the online Alexa Skills catalogue and given the option of downloading additional skills, for the following two weeks. The researcher visited the families again at the end of the study to ask them to reflect on their use of the Alexa skills.

## Data collection and analysis

Video recordings were taken during the first and second visits and conversations and interactions with Alexa transcribed. In addition, the voice recordings from when anyone interacted with Alexa were collected and transcribed. Both were coded separately for each visit by one rater using NVivo software, and all codings were reviewed by a second rater and placed into themes. These themes were extracted and validated as modelled by Saldana (2016), by discussing them with colleagues and checking interpretations with the participants themselves during the second visit.

## Findings

All the families tried out the three skills and several families downloaded other skills. There was much enthusiasm about how they had interacted with Alexa and how it became part of their family rituals. Usage of the skills persisted across the three weeks - as evidenced by the conversations recorded by Alexa. Some families reported using the device during meal times while others used it during shared activities such as cooking or after-school play. The younger children (aged 6) in the families were curious about Alexa, asking it a range of questions, as if it was a friend, for example, “*Alexa, what am I doing right now?*”, “*Hey Alexa, what's your second name and your third name?*”, “*Alexa, who is your favorite YouTuber?*”, “*Alexa, is Google and Siri your friends?*” One family (family 1) even resorted to asking Alexa during a meal time to resolve a debate: “*Once we used it when we were all eating, like we were having our meal and asking it questions... That's what it was, we were discussing something, and we didn't know who was right, so we asked Alexa...*”

To examine further the different ways Alexa became integrated into family life we identified and then categorized excerpts from the transcripts in terms of three main themes: family cohesion, family rituals and scaffolding. These were further classified in terms of the specific mechanisms used when interacting with Alexa. Certain themes recurred both within a single observation (i.e. behaviors were observed repetitively in the same family or a theme was mentioned a number of times within the same family) in addition to being a common theme across families. The quantity of incidences together with specific examples and comments made by the families are provided below. The most frequent interactions observed were in the scaffolding category, where different family members helped each other but especially younger children interact with Alexa.

### Family cohesion

Many forms of family cohesion was observed across all families. These were classified in terms of shared laughter, teasing, exchanging gestures and facial expressions and family rivalry.

*Shared laughter.* The interactions with Alexa often resulted in much shared laughter. 62 instances were identified during the three weeks for all families. They typically happened when family members were playing a game and when a mishap occurred, that Alexa was not privy to. For example, when playing the Kids Hub skill, family 3 chose the activity "Pictionary", a charades-style guessing game. In this game, Player 1 has ten seconds

to leave the room and following this Player 2 is given a word which they have to draw. Player 1 then comes back in the room, looks at what the other person has drawn and tries to guess the word. In one game, the son overheard Alexa give his mother the word "dog" to draw. They continued playing the game and the mother mentioned the word by mistake to which the son retorts:

Mother: *"This doesn't even look like a dog"*

Son: *"You just told me!"*

At which point they both laugh and the mother continues, *"Well you heard it anyway. I'm not good <at drawing> this looks like a cat, maybe even a pig"*

Shared laughter was also observed in situations where Alexa didn't work as intended, for example when it was unintentionally activated by a TV ad (family 4). Alexa's misunderstandings of what someone said were also a source of amusement. One child recalled a situation where Alexa misunderstood his mother's request: *"Oh, when you were asking how long to cook sausages for. And then it tried to order sausages!"* (family 3).

*Teasing.* There were also several instances of teasing (34) recorded, although considerably less than the shared laughter, across five of the families. Typically, family members interacted with Alexa to make fun of or provoke other family members in a playful way. For example, the child in family 2 was playing a music quiz online and losing to an opponent (Tom) in Texas. He kept looking at his dad to see if he knew the answers:

<Alexa plays next song>

<Child laughs and grimaces to his dad to indicate that he doesn't know the song>

Child: *"Skip"*

Alexa: *"The song is Timber by Pitbull, Tom answered correctly"*

Father: *"Tom needs to get out more."*

<Next song comes on and child looks at his dad and laughs>

Father: *"don't ask me!"*

The parents in family 1 often asked Alexa to play the good morning and the good evening song. When asked about this, the mother laughed and said that she would put the song on to annoy her son in the morning while he was getting ready for school or when it was time to get ready for bed.

*Exchanging gestures and facial expressions.* Families often reacted to interactions with Alexa by exchanging facial expressions, mimicking Alexa or making faces in response to things that it said. There were 41 instances of this coded for families 1,3,4 and 6. An example of exchanging facial expressions was when the mother and son (family 1) were listening to "fun facts":

Alexa: *"In the 18th Century it was believed that kissing a donkey could relieve toothache."*

<Son and mother make a disgusted face at each other >

An example of mimicking Alexa was when one family switched Alexa from having a British accent to an Australian one. The father then started speaking in an Australian accent which the children found funny. They then also tried to talk in an Australian accent. This led to much banter.

*Family rivalry.* A normal aspect of family life is sibling rivalry - when children think they are getting unequal amounts of attention or responsiveness from their parents, or intergenerational conflict. 45 instances of this were coded for families 2, 3, 4 and 5. Family 2 said they sometimes struggled to take turns to choose music because they did not like each other's choices. In contrast to other personal technologies, where the person holding a device or remote control has sole control, everyone has equal access to control of Alexa. This can mean anyone can override or interrupt what the other has asked Alexa to do as one father commented: *"It was like one of us would put music on and then the other one would change it when they couldn't put up with it anymore, so we weren't taking turns."* One mother also explained how when her children try to use Alexa at the same time it could lead to arguments. For example, the siblings in Family 4 often competed for Alexa's attention with different requests and then became upset when the mother gave her support to the other one:

Mother: *"Alexa play barking"*  
Youngest child: *"I don't want that!"*  
Mother: *"Well you're gonna get it."*  
Youngest child: *But he got what he wanted!*

Another time a mother (family 5) tried to reason with her elder child about needing to be fair with his younger brother:

Mother: *"Did you take it in turns?"* <when asking Alexa for music>  
Elder child: *"Yeah, but then he got to choose."*  
Mother: *"Well probably because you'd been up for an hour already and had loads of choices before he came and had breakfast."*

### Family rituals

Family rituals are an integral part of family life. These were much in evidence when interacting with Alexa, with 58 instances coded across all families. Most were triggered during the playing of a song. For example, when a child requested Alexa play the 'Baby Shark' song every day, the mother groaned each time, but then the whole family would start dancing and singing along. Another example, is where the mother of family 1 requested a song on her son's behalf to which the father looked at the son and encouraged him to show off his dance moves:

Mother: *"Alexa play the gummy bear song."*  
<Son looks at his mum>  
Father: *"Show her your dance moves"*

Another child (family 4) commented how their parents now refer them to ask Alexa if they want to find something out rather than looking it up themselves on the internet: *"If mum or dad are doing something and I ask them a question instead of going like have you searched it up they would say go and ask Alexa."* This shift requires them formulating and talking aloud their query which the parents can be privy to if in the same room and join in or be aware of what their children are trying to find out.

### Scaffolding

A common occurrence was when family members collaborated by encouraging each other when using Alexa. 111 instances of these were observed across all families – making this the most frequently observed category. This happened either when the family worked together as a team or when a more competent family member helped a younger member interact with Alexa.

When playing a story skill with Alexa, such as PacMan, family members sometimes whispered to each other what they should do before asking Alexa – so as not to let it hear or know about their discussion before making their choice. Thumbs up gestures and nodding were also employed during this kind of behind the scenes negotiation process. Sometimes one person would take the role of leader while the rest of the family members would whisper to each other what to answer. For example, family 4 discussed which option to choose before the youngest child eventually shouted out their answer:

Alexa: *"You could pick up the fork or you could read the sign."*  
<All whisper to each other to decide which one.>  
Mother: *"What do you want to do?"*  
Youngest child shouts: *"Pick up the fork!"*

When Alexa did not recognize what a younger member of the family said, other family members often repeated it slowly to encourage the child to try again but to speak more clearly. For example, the youngest child in family 4 wanted to play a funny song but was not able to say the word of the song correctly. His older brother helped him by repeating it and gesturing at Alexa to indicate where to speak. One mother (Family 5) commented on how it was much harder for the younger child to be understood by Alexa. When they struggled with some of the questions she would try to rephrase it to be simpler, so the child could learn how to be more concise. Similarly,

in family 2, the father pointed out how Alexa does not understand his 11-year-old son as well as himself. Another mother (Family 1) often encouraged her child to interact with Alexa by pointing at it:

Mother to child: *Do you want to hear another?*

<Mother points at the Alexa to encourage child to ask>

Child to mother: “yes”

Another strategy families used to encourage their children to speak to Alexa was to tell them how to ask Alexa questions. Families also encouraged their children to practice speaking to it. Hence, there were many examples of parents helping their children to learn how to speak with Alexa to make themselves understood.

## Discussion

The findings demonstrated how families engaged in a diversity of interactions when using the Alexa skills. They developed new family rituals and encouraged each other to talk to and take part in games with Alexa. Many of these interactions could be seen as contributing to social and emotional bonding, leading to further family cohesion. For example, the use of non-verbal behavior while interacting with Alexa provided a means of expressing empathy. They all took to having a new voice in the family home – like having a new family pet. However, the interactions that occurred were not always harmonious and sometimes led to family disputes. In a way, this reflects the richness of family life, where children learn to compete with their siblings but also learn to empathize with them; where families develop new rituals by playing games or performing together and in doing so bond over them but at other times use the opportunity to tease each other.

Lots of shared laughter occurred because of Alexa’s mishaps especially when it misunderstood them. For younger children, this could be quite frustrating and at times challenging. Parents and elder children were acutely aware of what they were going through and would often try to help them out, by showing them how to interact with Alexa. Other times, activities with Alexa led to joking, joshing and teasing, building upon pre-existent family discourses about particular interests or sensitivities. Alexa was also incorporated into existing family dynamics and could sometimes encourage or amplify those, for example, when commenting on different tastes in music or different realms of expertise across family members. Different interests across family members could also lead to disagreement, particularly when children wanted to play certain songs or games repeatedly. This happened both between siblings and between parents and children.

Alexa in its current form was unable to recognize different family members by their voice. This meant it could not understand different people talking simultaneously – which families often do in family conversations, and during which they are able to hone in on what one person is saying (cf. the cocktail party effect). To adapt to Alexa not being able to do this, families adapted their behavior. For example, they would all sit together, often on the same sofa, facing Alexa and communicate with it, as a family unit, rather than carry on having a conversation with each other. Alexa became the central focus for their interaction. This differs from Porcheron *et al.*’s (2018) finding where the conversation with Alexa was interwoven with the parents taking turns to talk to cajole their child to eat their dinner.

Families frequently used back-channel methods of communicating, such as whispers, to avoid getting unintended responses from Alexa. Sometimes this was done in the spirit of pretend play, where families whispered together so that the children would learn how to make joint decisions together when playing a game that others would not be able to hear – in this case Alexa. Once agreed on their answer, an elected person (self or other) in the family would call it out to Alexa. This pattern of interaction was observed when families played games together. However, problems sometimes occurred when there were younger children in the family who found it difficult to follow the rules of playing a game or a quiz, or when siblings felt unequally treated. When such times arose, families changed tack to deal with them, for example, the parents and elder children adopted helper roles to encourage and make suggestions for what the younger child should say to Alexa.

Alexa was also found to encourage families to jointly act out together, namely singing and dancing. Having an external participant made this appear like a performance, as evidenced by parents sometimes referring to Alexa as a potential audience, for example, suggesting that their children show Alexa their dance moves. Listening to music together is something families may already do, for example on journeys in the car. Teenagers may listen to music together and act out the singer’s dance routines. With Alexa, it became possible for families to feel comfortable with all the members dancing and singing repeatedly. Even the mother groaning appeared to enjoy being part of the ritual. And, as Druga *et al.*, (2017) point out, family rituals have a positive impact on family cohesion and emotional well-being. In terms of learning, children were inducted into a whole range of skills in how to tailor their communication to a new audience member, how to hold the floor, and how

to use back channels of communication when there are interactants with different needs. Having a new interactant in the house also gave new possibilities for performance, and engagement with music in particular. Further work will reveal how the role of voice assistants in the family may evolve over the longer term.

## Conclusion

Voice assistants, like Alexa, are making big in-roads into family homes. Our study showed how families readily appropriate them into their family life, playing, singing, performing and even asking Alexa questions to resolve family disputes. Alexa was found to encourage much shared laughter and helping each other but also trigger sibling rivalry, teasing and family arguments. However, both negative and positive aspects of family interactions are an integral part of learning about each other and developing empathy skills. Contrary to Turkle's (2015) concerns about the digital age, interactions with voice assistants such as Alexa, may even help children learn the art and joy of conversation in the presence of humans and machines.

## References

- Bhömer, M.T., Helmes, J., O'Hara, K. & van den Hoven, E. (2010). 4Photos: a collaborative photo sharing experience. In *Proc. of NordiCHI '10*. ACM Press, 52-61.
- Boer, D., & Abubakar, A. (2014). Music listening in families and peer groups: Benefits for young people's social cohesion and emotional well-being across four cultures. *Frontiers in Psychology*. 5.
- Davidoff, S., Lee, M.K., Yiu, C., Zimmerman, J. & Dey, A. (2006). Principles of smart home control. In *International conference on ubiquitous computing (UBICOMP)*. Springer, Berlin, Heidelberg. 19-34
- Doucleff, M. and Aubrey, A. (2017). Alexa, Are You Safe for My Children? <https://www.npr.org/sections/health-shots/2017/10/30/559863326/alexa-are-you-safe-for-my-children>.
- Druga, S., Williams, R., Breazeal, C. & Resnick, M. (2017). "Hey Google is it OK if I eat you?" In *Proc. of the 2017 Conference on Interaction Design and Children (IDC '17)*, ACM Press. 595–600.
- Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems*. 42(3-4), 143-166.
- Ganesh, S., Marshall, P., Rogers, Y. & O'Hara, K. (2014). FoodWorks: tackling fussy eating by digitally augmenting children's meals. In *Proc. of NordiCHI'14*. ACM Press, 147-156.
- Kaye, J., Fischer, J., Hong, J., Bentley, F.R., Munteanu, C., Hiniker, A., Tsai, J.Y. & Ammari, T. (2018). Panel: Voice Assistants, UX Design and Research. In *Proc. Conference on Human Factors in Computing Systems (CHI EA '18)*. ACM, New York, 01, 5 pages.
- Nass, C., Moon, Y., & Carney, P. (1999). Are people polite to computers? Responses to computer-based interviewing systems. *J Appl Soc Psychol* 29, 5: 1093-1109.
- Plowman, L., Stephen, C. & McPake, J. (2010). *Growing up with technology: Young children learning in a digital world*. New York: Routledge.
- Plowman, L., Stevenson, O., Stephen, C. & McPake, J. (2012). Preschool children's learning with technology at home. *Computers & Education*, 59 (1), 30–37.
- Porcheron, M., Fischer, J.E., Reeves, S. & Sharples, S. (2018). Voice Interfaces in Everyday Life. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. 640, 12 pages.
- Purinton, A., Taft, J.G., Sannon, S., Bazarova, N.N. & Taylor, S.H. (2017). "Alexa is my new BFF": Social Roles, User Satisfaction, and Personification of the Amazon Echo. *Proc. of the CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '17)*. ACM Press. 2853–2859.
- Reeves, B. & Nass, C. (2006). *The media equation: How people treat computers, television, and new media like real people and places*. Stanford, CA: CSLI.
- Saldana, J. (2016). *The coding manual for qualitative researchers*. Sage.
- Sciuto, A., Saini, A., Forlizzi, J. & Hong, J. (2018). "Hey Alexa, What's Up?: A Mixed-Methods Studies of In-Home Conversational Agent Usage." In *Proceedings of DIS'2018*, 857-868. ACM, NY.
- Short, E.S., Swift-Spong, K., Shim, H., Wisniewski, K.M., Zak, D.K., Wu, S. & Matarić, M.J. (2017). Understanding social interactions with socially assistive robotics in intergenerational family groups. *26th IEEE International Symposium on Robot and Human Interactive Communication*. 236-241.
- Turkle, S. (2015). *Reclaiming conversation: the power of talk in a digital age*. Penguin.
- Wang, B., Taylor, L. & Sun, Q. (2018). *Families that play together stay together: Investigating family bonding through video games*. New Media & Society.
- Weisner, T. (2002). Ecocultural understanding of children's developmental pathways. *Hum. Dev.*, 45, 275-281.
- Yuill, N. & Martin, A. (2016). Curling up with a good E-Book: Mother-child shared story reading on screen or paper affects embodied interaction and warmth." *Frontiers in psychology* 7 (2016): 1951.

# Measuring the Computational in Computational Participation: Debugging Interactive Stories in Middle School Computer Science

Chris Proctor, Stanford Graduate School of Education, cproctor@stanford.edu

**Abstract:** An equitable implementation of K-12 computer science must support inclusive literacy practices, but it must also develop concrete skills. This study analyzes the extent to which a computer science curriculum based on digital storytelling helped students become more effective at debugging. Prior research has developed digital storytelling as a medium for computational participation, but few studies have reported detailed results on growth in computer science skills. Meanwhile, research on debugging has tended not to address sociocultural factors. This study, conducted over four months of a middle-school computer science course using interactive storytelling, analyzed student reading, writing, and debugging practices based on approximately 1000 story edits and user behavior collected from the platform's logs. The results suggest that literacy-based computer science education using digital storytelling can be productive for developing skills such as debugging.

## Introduction

Driven by the economic opportunities and pervasive societal impacts of computing, computer science (CS) is rapidly becoming a mainstream subject in primary and secondary education. Computational thinking is developing as a set of ideas and practices to be taught across the curriculum. In either case, the interdisciplinary connections to STEM subjects are clear and compelling (Weintrop, et al., 2014), while much less is said about how CS might support the core concerns of the humanities and social sciences. Indeed, when CS is defined narrowly as a collection of facts and skills about solving problems with computers, the subjects may not have much common concern. However, youth today engage in diverse and complex literacy practices with digital media (Ito, et al., 2010), relying to various extents on the computational aspects of these media to engage in computational participation (Burke, O'Byrne, & Kafai, 2016). If K-12 CS were contextualized within these literacies, its central concepts and practices could become an essential part of diverse youth cultures reading, writing, and analyzing digital texts.

Recognizing this opportunity (particularly for youth who do not see their identities and cultures represented in the world of CS), there have been numerous efforts to incorporate digital storytelling into school. Like projects which embed computation into animation (Resnick, et al., 2009) and e-textiles (Buechley, 2006), digital storytelling allows learners to learn programming and encounter powerful ideas from computer science using media which already mediate their literacy practices. Ware & Warschauer (2005) explored the potential for digital storytelling to transcend the divide between school and informal spaces for youth marginalized by race and social class. Kelleher & Pausch (2007) conducted digital storytelling workshops with middle-school girls using a modified version of Alice, finding that the opportunity for self-expression, sharing, and identity development provided motivation for learning to program. Burke & Kafai (2010) drew on parallels between programming and writing to explore how digital storytelling might support growth in each. Proctor and Blikstein (2019) analyzed how the computational elements of interactive stories could function rhetorically and support the development of critical perspectives.

However, some computer scientists argue that sociocultural definitions of K-12 computer science, such as those above, are unworkably vague and impossible to assess (Denning, 2017). This critique demands a response, particularly as it aligns with broader arguments that specific skills are best taught in a context that minimizes cognitive load (Kirschner, Sweller, & Clark, 2006). Must the literacy-based approaches described above involve a tradeoff in terms of how well students learn foundational skills and concepts? Building on prior work developing digital storytelling as a fruitful medium for computational literacies, this study analyzes whether such an environment can also effectively support specific computational practices such as debugging. This study's research questions are:

1. Does writing interactive stories support development of debugging practices?
2. Does reading other stories support development of debugging practices?

## Background

## Text-based interactive storytelling

Interactive storytelling can be distinguished from the broader category of digital storytelling by the use of programming to implement nonlinear single-player games in which the player becomes a character. Text-based interactive stories are particularly effective settings for using computational elements for rhetorical effect in the service of representing and critically analyzing social realities (Proctor & Blikstein, 2019; Proctor & Garcia, 2019). In contrast to primarily-visual storytelling platforms such as Scratch and Alice, writing is singularly important for narrative, representation, and analysis of subjective phenomena in popular culture and in the humanities and social sciences. One example of the transmedia possibilities of text-based interactive storytelling is a high school sociology class in which students wrote interactive stories exploring the use of power in social interactions. The stories' use of programming allowed them to create replayable models of social situations in which readers could explore the consequences of different interactional choices.

The implementation used in this study is a web application called Unfold Studio in which stories' source code is presented side-by-side with a running version. Unfold Studio has social affordances such as the ability to

```

Welcome to California
+ I am a 14 year old girl.
  -> student
+ I am a 50 year old man.
  -> teacher

=== teacher ===
As an adult you have special
priviledges.
+ Fly to San Fransisco
  -> enter

=== student ===
You can show your parent you
can do something on your own!
+ Fly to San Fransisco for a
short independace trip.
  -> enter

=== enter ===
You made it to San
Fransisco!!!
+ Go to the hotel
  -> hotel

=== hotel ===
{teacher > 0:
  The Man at the front
desk says, "Hello Sir, do you
have a reservation for a
room?"
  -> room
- else:
  The Man at the front
desk says, "Why are you here?
Your dont have a reservation!
Where are your parents? Go
away child!" Oh no! Your
stuck to sleep on the steets!
  -> streets
}
```

Figure 1. Excerpt from an interactive story written in the Ink programming language.

publish stories, browse and read other authors' stories, and a feed showing events related to an author's stories and other authors she follows. Stories are written in a programming language called Ink (Inkle, 2016) and compiled every time an author saves her work. The story excerpt shown in Figure 1 illustrates the syntactical elements analyzed in this study. Chunks of the story are defined as knots (=== hotel ===), which are linked together via diverts (->). Knots typically end by presenting the player with choices (+) of what should happen next. Constructs within curly braces allow variables to influence the content and choices shown to the player. When stories contain errors, explanatory error messages are shown in the space that would have been taken up by the running story.

## Debugging

This study analyzes the association between writing interactive stories and performance on a debugging task. McCauley, et al.'s (2008) review of educational literature on debugging defines debugging as part of a response to some kind of breakdown in a programmer's plan for reaching a goal. After testing reveals that a breakdown has occurred, debugging involves "find[ing] out exactly where the error is and how to fix it." (p. 68) The K-12 Computer Science Framework (2016) uses a similar definition and adopts testing and debugging as one of seven core practices in computer science (p. 81). Although debugging has historically not been emphasized in computing education, it occupies a substantial portion of professional programmers' time (Beller, et al., 2017). Bugs may occur at the level of syntax (the program cannot be compiled), semantics (the program crashes at runtime), or logic (the program works, but not as expected). The available data limits this study to an analysis of bugs in syntax.

Debugging is a particularly interesting skill to study in an interactive storytelling context. On the one hand, prior research has found debugging to be a distributed social practice, which suggests a literacy-oriented approach could be effective in developing students' debugging skills. Flood, et al. (2018) found that learning to debug resembles enskilment in which "a newcomer is supported in appreciating and using the affordances of their environment" (p.1). Piorkowski, et al. (2013), found that during debugging, professional programmers spend half their time in information-foraging behaviors such as reading other programs, reading documentation, and searching online. Multiple studies have found comprehension of the program being debugged (and understanding of its goal) to be an important factor in debugging



success (McCauley, et al., 2008). This invites comparisons to the importance of reading comprehension for fixing grammatical errors in writing (Weaver, 1996).

On the other hand, foundational research on debugging identified misconceptions which could be exacerbated by a literacy-based approach. Bonar and Soloway (1985) found that novices misunderstand programming as writing. In particular, novices tended to inappropriately apply natural language meanings to words such as "while." Pea (1986) saw this as an instance of a more general superbug: novices' misconception that computers reason about programs and interpret them intelligently, instead of following them mechanically. It is plausible that text-based interactive storytelling, in which authors blend prose with code to implement discourse scenarios, could unproductively mingle the ways humans interpret text and the ways machines interpret code. However, it is also plausible that these misconceptions could actually be ameliorated by casting programming as reading and writing. When there is real authorial meaning motivating the program, and a real audience to whom it is addressed, it could be the case that students better understand the computer's mediating role, rather than misunderstanding it as co-author or interlocutor.

## Methods

### Context

The study took place over four months in a private all-girls middle school in the western United States. The participants, in seventh grade, were enrolled in a computer science class which met twice weekly for a 90-minute block period. 40 students out of a cohort of 67 consented to participate in the study. Students were asked to self-identify with respect to race and gender. 92.5% affirmatively identified as female; 37.5% identified as white, 27.5% as of mixed race, 20% as Asian, 5% as hispanic, and 10% declined to state. While no data on socioeconomic status were available, the school is located in an affluent area and charges substantial tuition while also offering full or partial scholarships to many of its students.

During the period of the study, the students' computer science class followed the curriculum developed by Proctor & Blikstein (2019). Students were introduced to increasingly complex computational concepts and programming syntax and asked to use them in open-ended story prompts. During class time, students were free to work on their stories, read their peers' stories, and to seek help from peers or the teacher. The study's author worked with the teacher to plan the unit and provided technical support, but was not involved in teaching.

### Data sources

#### Stories

There are 327 stories from 39 authors with an average of 92 lines ( $\sigma = 84$ ) and 431 tokens ( $\sigma = 400$ ), where tokens include individual words as well as syntactical elements. A snapshot was captured every time a student saved her story. Saving a story was a necessary step in recompiling and re-playing the story. Figure 2 (b) shows that stories generally grew quickly at first, and then entered a period of fluctuating slow growth. 83% of story edits took place within the first 90 minutes (the duration of a block class period) of the story's creation, so most of the activity under investigation in this study took place within the classroom literacy space.

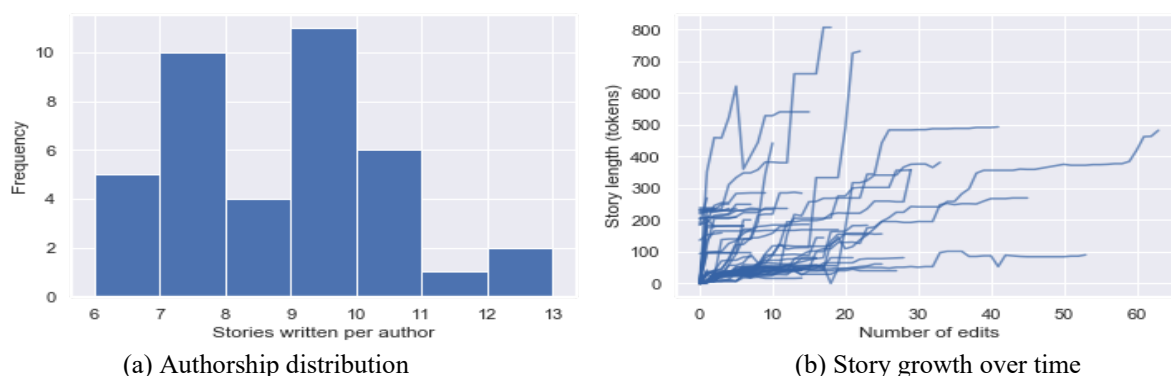


Figure 2. Stories written during the study.

This study's primary unit of analysis is the story edit, or differences between each pair of adjacent versions of a story. Comparing line-by-line changes in an edit yields a collection of ops, where each op is a contiguous set of lines inserted, deleted, or replaced from one version to the next. Each op was classified by the syntax elements

involved, such as knots or diverts. Some ops involve changes to code, such as creating new knots, diverts, and variables. Others only involve changes to text, leaving the story's programmatic structure unchanged. Then each edit was classified according to the properties of its ops and whether the prior and latter versions successfully compiled. Edits leaving the story unchanged were filtered out. Story edits inserted an average of 5.3 lines per edit ( $\sigma = 11.4$ ) and deleted an average of 1.9 lines per edit ( $\sigma = 4.2$ ). Table 1 shows the classification scheme.

Table 1: Classification scheme for states of story edits

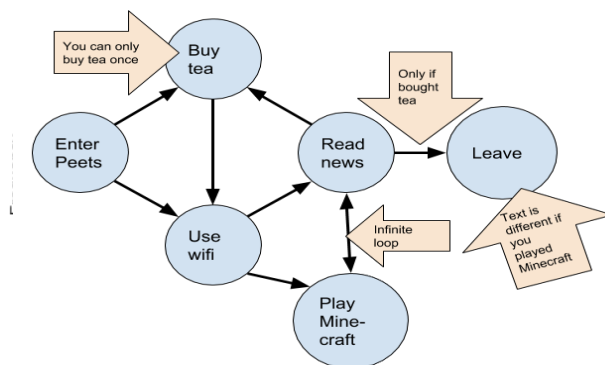
| Pre ok? | Post ok? | Ops                                       | State           | Simple state | Percent |
|---------|----------|---|-----------------|--------------|---------|
| yes     | yes      | $\geq 5$ lines inserted total             | MAJOR INSERT    | OK           | 14%     |
| yes     | yes      | $-5 < \text{lines} < 5$ ; code changes    | MINOR CODE EDIT | OK           | 27%     |
| yes     | yes      | $-5 < \text{lines} < 5$ ; no code changes | MINOR TEXT EDIT | OK           | 22%     |
| yes     | yes      | $\geq 5$ lines deleted total              | MAJOR DELETE    | OK           | 2%      |
| no      | yes      | any                                       | DEBUG SUCCESS   | OK           | 10%     |
| yes     | no       | any                                       | ERROR           | ERROR        | 11%     |
| no      | no       | any                                       | DEBUG FAIL      | ERROR        | 13%     |

### Views of other stories

The second research question asks whether reading other stories was helpful in students' debugging. To answer this, backend log data was filtered to collect every instance of a student viewing a story. We then counted the number of a user's views which occurred during the span of each story edit (excluding views of the story being edited). Students viewed an average of 2.4 other stories ( $\sigma = 2.6$ ) during editing. This was surprising, as with only a few exceptions, students reported in the post survey that they enjoyed reading each others' stories and frequently cited specific examples of stories they had read. One possible explanation is that this study counts only views that took place during an edit.

### Summative assessment and survey

Two months after the interactive storytelling unit ended, participants were given a summative assessment of their debugging skills and a post-study survey collecting demographic information and measuring their affect and attitudes toward computing (Friend, 2016). Because participants had no prior experience with interactive storytelling, a pre/post test study design was not feasible. We used an adaptation of the Fairy performance assessment (Werner, et al., 2012), which was designed for this situation. Students were given a directed graph (Figure 3(b)) showing the desired functionality of a story (they had previously used directed graphs in planning their stories), annotated with specific issues. They were then given a copy of a broken implementation of the story and asked to fix it.



(a) Score distribution (b)

Story graph showing desired functionality

Figure 3. In the summative assessment, students were asked to debug a broken story.

These assessments were scored using a rubric. One point was assigned for correctly linking story knots together using diverts (mapping the graph structure to the story structure), and one point was assigned for correcting each issue pointed out by an arrow. Each required the use of a different computational topic students had worked with during the unit: maintaining

state explicitly with variables and implicitly with the players' path through the story, using state to change text output, and using state to control the availability of choices.

## Analysis

RQ1: Does writing interactive stories support development of debugging practices?

To answer this question, we first needed to describe patterns in students' story editing. We created a matrix counting the number of transitions from each pair of states, across all edits in all stories, and visualized this in a transition diagram. We also created a simplified model grouping together all successful edits and all unsuccessful edits. We then used OLS regression to estimate the association between writing more complex stories and summative score, and the association between probability of successful debugging and summative score. Story complexity was defined as the sum of diverts and choices used across all of an author's stories. This measure was chosen to capture the amount of computational content in stories. The probability of successful debugging was calculated as  $DEBUG\_SUCCESS / (DEBUG\_SUCCESS + DEBUG\_FAIL)$  for each author.

## RQ2: Does reading other stories support development of debugging practices?

We theorized that as debugging is a distributed, social, and mediated process, students might be especially likely to view other stories (either written by peers or their own earlier stories) during debugging. As described in the previous section, we collected the number of other story views which occurred during each story edit. We grouped edits into debugging edits ( $DEBUG\_SUCCESS$  and  $DEBUG\_FAIL$ ) and non-debugging edits (the rest), and conducted a two-tailed, two-sample independent T-test to determine whether a significant difference exists in these two groups' means, with the null hypothesis that there is no difference between group means. We do not assume equal sample size or variance between the groups, so we use Welch's T-test. Additionally, in the post-survey survey, we asked students whether they found it helpful to look at other stories.

## Results

### RQ1: Does writing interactive stories support development of debugging practices?

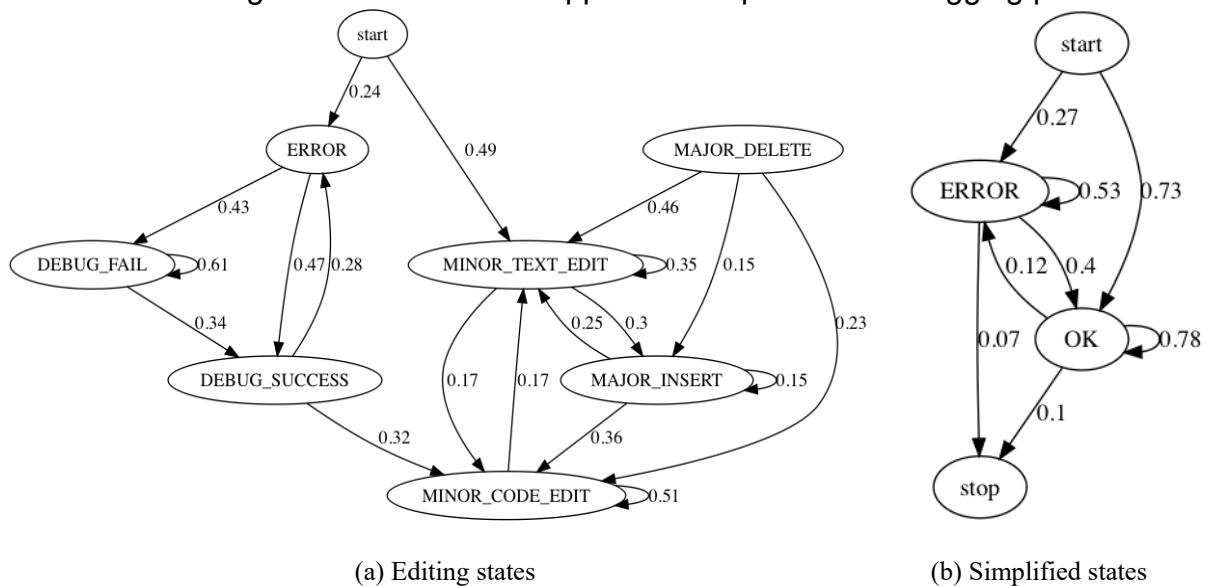


Figure 4. Transition diagrams showing probabilities of moving between story editing states.

Figure 4 shows the full (a) and simplified (b) transition diagrams between story editing states. (Transition probabilities under 0.15 are omitted for clarity.) Several editing patterns are visible: Authors tend to begin by entering one of two subgraphs: success or error (this was the inspiration for creating the simplified diagram in Figure 4(b)). When edits are successful, authors generally begin with short text editing (for example, writing a story with no code), and then cycle between editing text and code.  $MINOR\_CODE\_EDIT$  is a sink state, suggesting that once stories successfully mature, authors tend to spend their effort tweaking the code rather than adding substantial new textual content.

Of the successful editing states,  $DEBUG\_SUCCESS$  stood out as most likely to precede an error. This is not surprising, as creating and fixing errors is part of the iterative process of writing challenging programs. When authors fail to successfully debug on encountering an error, they enter the  $DEBUG\_FAIL$  sink state. The overall impression given by Figure 4(a), and clarified in Figure 4(b), is that of generally productive editing with distinct modes of successful editing and debugging.

Figure 5(a) shows a scatter plot and regression line (with 95% confidence interval) for the association between an author's total output of diverts and choices and score on the summative assessment. After removing the prolific outlier (shown in orange), the association was modeled as  $\text{score} = 1.58 + (\text{diverts} + \text{choices}) * 0.0035$ , with  $P > |t| = 0.004$  and adjusted  $r^2 = 0.18$ . In contrast to this strong association, there was no association between probability of successful debugging and summative score. This is clearly visible in Figure 5(b).

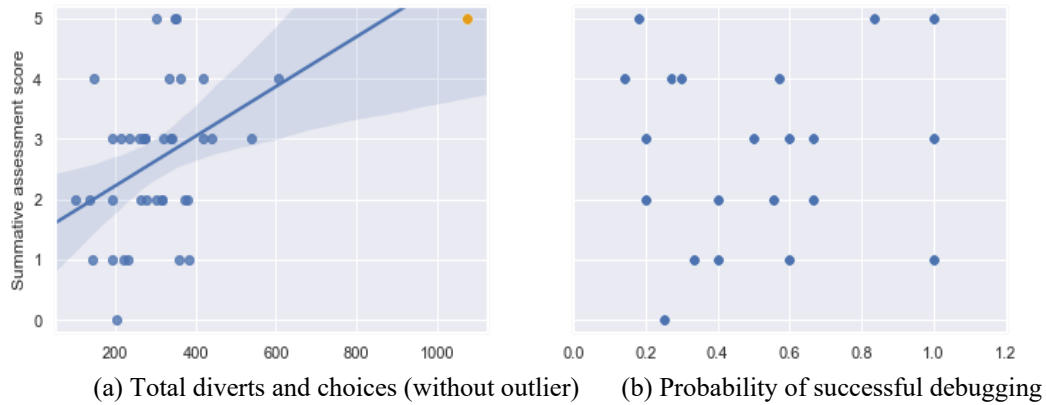


Figure 5. OLS Regression and 95% confidence interval plotted.

## RQ2: Does reading other stories support development of debugging practices?

Figure 6(a) shows the mean and standard deviation number of story views for each editing state. The result of the T-test comparing debugging states and others was  $t = -3.66$ ,  $p > |t| = 0.0003$ . There was a significant difference between debugging and non-debugging states, but in the opposite direction as was hypothesized. Another unexpected difference was the dramatically higher number of story views during MINOR TEXT EDIT and ERROR states. We wondered whether this was due to these states tending to have longer time durations; Figure 6(b) plots the mean time interval for each editing state. We interpret these results in the next section.

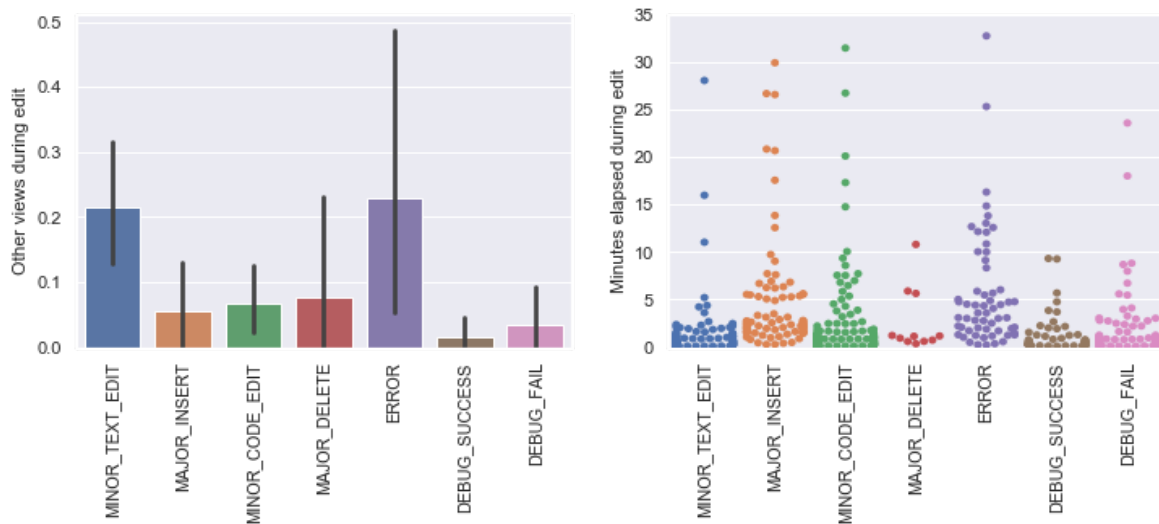


Figure 6. Characteristics of different edit states.

## Discussion

This study's results suggest that a literacy-based approach to teaching computer science using interactive storytelling can be an effective context for learning debugging. The results also revealed surprising dynamics between debugging and viewing other stories, which warrant further iterations of design-based research. Participants in this study engaged in substantial debugging (moving in and out of error states) during their story-writing, and authors who used more computational elements in their stories were more likely to score highly on

the summative assessment. However, we did not find an association between successful debugging and higher performance on the summative task. These findings agree with participants' near-universal consensus that the literacy-based approach to teaching computer science helped them grow as programmers. Furthermore, when the post-study survey was administered, students had already spent two months learning Python, and 89% felt that interactive storytelling with Ink had helped prepare them to learn Python. (In contrast, transitioning from block-based languages to text-based languages is often difficult for novices (Weintrop & Wilensky, 2016).

One surprising finding was that, rather than using other stories as debugging resources, authors viewed stories less often during debugging. One interpretation of this is might be that debugging is a time to focus on the problem. Intuitively there is some truth to this, however this would also imply that looking at other stories during debugging should be negatively-correlated with debugging success. We did not find this to be the case. Furthermore, the literature on professional debugging (discussed above) finds a substantial reliance on other people and texts. It could be the case that other resources, such as handouts or the Ink language documentation, were more useful than other stories when authors got stuck.

The other unexpected pattern observed in story views across edit states was that authors looked at other stories so frequently while in ERROR and MINOR TEXT EDIT. The prevalence of other story views during ERROR may be partially explained by the fact that ERROR states tended to occupy somewhat more time than other states (see Figure 6(b)). Additionally, it is plausible that authors in ERROR edit states became frustrated and disengaged, and read other stories as a diversion rather than as part of productive debugging. The disproportionately high views of MINOR TEXT EDIT, however, cannot be explained by long duration; 90% of these edits spanned less than three minutes. The post-study survey provides a possible interpretation. All but three participants felt that reading peers' stories was helpful, but when asked to cite specific examples, students almost always described the content of the stories, and not their structural features. This, together with the high number of other views for the MINOR TEXT EDIT state, suggests that students may have been more successful learning from the content of each others' stories than from the computational aspects.

If this interpretation is correct, it poses a design challenge for future development: How might we design the literacy space so that students can learn from computational aspects of other stories as well as their content? One possible approach was tested at the end of the study, and has since been developed further. The students' final assignment was to collaboratively create a story in which the reader explores a world. The class planned out the story on one wall of the classroom, each student wrote a small part of the world, and then they worked together to import partial stories into a whole. Variables (such as energy level or the number of clues collected) maintained state across the different components. This assignment surfaced important computational concepts such as state and interface, and made students dependent on the computational aspects of each others' stories.

The results of this study should not be over-generalized. This study's participants attended an all-girls private school which provides unusual levels of personal attention (possibly allowing them to feel safe participating in a literacy space) and they had already studied computer science for one year using Scratch. This study's quantitative results have yet to be replicated with larger and more heterogeneous participants, and in-person ethnographic analysis of interactions within the literacy space is needed to validate our interpretations. This research is underway.

## Conclusion

At the conclusion of CSCL 2017, eight provocations were presented for the future of the field, including the question of whether the community ought focus on basic research and give up trying to make tangible change in the educational system (Wise & Schwartz, 2017). The prospect of widespread adoption of computer science and interdisciplinary computational thinking potentially means that the debate will shift from whether computers should be used in schools, to how. This ought to reinvigorate sociocultural research on how schools can support vibrant human-computer activity systems which attend to questions of culture, identity, power, and privilege, as well as developing students' technical skills. This paper contributes to the proposition that we do not have to choose between those goals.

## References

- Beller, M., Spruit, N., & Zaidman, A. (n.d.). How developers debug.
- Bonar, J., & Soloway, E. (1985). Preprogramming knowledge: A major source of misconceptions in novice programmers. *Human-Computer Interaction*, 1(2), 133-161.
- Buechley, L. (2006). A Construction Kit for Electronic Textiles. In *2006 10th IEEE International Symposium on Wearable Computers* (pp. 83-90). Montreux, Switzerland: IEEE.
- Burke, Q., & Kafai, Y. B. (2010). Programming & storytelling: opportunities for learning about coding & composition (p. 348). ACM Press.

- Burke, Q., O'Byrne, W. I., & Kafai, Y. B. (2016). Computational Participation: Understanding Coding as an Extension of Literacy Instruction. *Journal of Adolescent & Adult Literacy*, 59(4), 371–375.
- Denning, P. J. (2017). Remaining trouble spots with computational thinking. *Communications of the ACM*, 60(6), 33–39.
- Flood, V. J., DeLiema, D., Harrer, B. W., & Abrahamson, D. (2018). Enskilment in the Digital Age: The Interactional Work of Learning to Debug, 2.
- Friend, M. (2016). *Women's History, Attitudes, and Experiences with Computing*. Dissertation.
- Inkle. (2016). *Ink*. London, UK. Retrieved from <https://github.com/inkle/ink>
- Ito, M. (2010). *Hanging out, messing around, and geeking out: Kids living and learning with new media*. Cambridge, MA: MIT Press.
- K-12 Computer Science Framework. (2016).
- Kelleher, C., & Pausch, R. (2007). Using storytelling to motivate programming. *Communications of the ACM*, 50(7), 58.
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist*, 41(2), 75–86.
- McCauley, R., Fitzgerald, S., Lewandowski, G., Murphy, L., Simon, B., Thomas, L., & Zander, C. (2008). Debugging: a review of the literature from an educational perspective. *Computer Science Education*, 18(2), 67–92.
- Pea, R. (1986). Language-independent conceptual bugs in novice programming. *Journal of Educational Computing Research*, 21, 25–36.
- Piorkowski, D. J., Fleming, S. D., Kwan, I., Burnett, M. M., Scaffidi, C., Bellamy, R. K. E., & Jordahl, J. (2013). The whats and hows of programmers' foraging diets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13* (p. 3063). Paris, France: ACM Press.
- Proctor, C., & Blikstein, P. (2019). Unfold Studio: Supporting critical literacies of text & code. *Information and Learning Science*, 2(1).
- Proctor, C., & Garcia, A. (2019). Hogg, L., & Stockbridge, K. (Eds.). Student voices in the digital hubbub. *Giving student voice due weight: Possibilities and challenges in USA and New Zealand*.
- Resnick, M., Maloney, J., Monroy-Hernández, A., Rusk, N., Eastmond, E., Brennan, K., ... others. (2009). Scratch: programming for all. *Communications of the ACM*, 52(11), 60–67.
- Ware, P. D., & Warschauer, M. (2005). Hybrid literacy texts and practices in technology-intensive environments. *International Journal of Educational Research*, 43(7–8), 432–445.
- Weaver, C. (1996). *Teaching Grammar in Context*. Portsmouth, NH: Boynton/Cook.
- Weintrop, D., Beheshti, E., Horn, M., Orton, K., Jona, K., Trouille, L., & Wilensky, U. (2016). Defining Computational Thinking for Mathematics and Science Classrooms. *Journal of Science Education and Technology*, 25(1), 127–147.
- Weintrop, D., & Wilensky, U. (2016). Bringing Blocks-based Programming into High School Computer Science Classrooms, AERA.
- Werner, L., Denner, J., Campe, S., & Kawamoto, D. C. (n.d.). The fairy performance assessment: measuring Cauleycomputational thinking in middle school, 6.
- Wise, A. F., & Schwarz, B. B. (2017). Visions of CSCL: eight provocations for the future of the field. *International Journal of Computer-Supported Collaborative Learning*, 12(4), 423–467.

# Does Order Matter?

## Investigating Sequential and Cotemporal Models of Collaboration

Zachari Swiecki, University of Wisconsin—Madison, swiecki@wisc.edu

Zheming Lian, University of Wisconsin—Madison, zlian4@wisc.edu

Andrew R. Ruis, University of Wisconsin—Madison, arruis@wisc.edu

David Williamson Shaffer, University of Wisconsin—Madison and Aalborg University Copenhagen,  
dws@education.wisc.edu

**Abstract:** Many researchers have argued that models of collaborative processes should account for temporality, but there exist different approaches for doing so. We compared two specific approaches to modeling collaborative processes in a CSCL context: Epistemic Network Analysis, which models events cotemporally (unordered and temporally proximate), and Sequential Pattern Mining, which models events sequentially (ordered and temporally proximate). Our results suggest that in this context cotemporal models constructed with Epistemic Network Analysis outperform sequential models constructed with Sequential Pattern Mining in terms of (a) explanatory power, (b) efficiency, and (c) interpretability.

### Introduction

A central claim of computer-supported collaborative learning (CSCL) research is that collaborative processes influence group performance, and studies have shown that these processes have an important temporal dimension (Reimann, 2009). Researchers thus argue that models of collaborative processes should account for temporality (McGrath & Tschan, 2004), but there exist different approaches for doing so. *Sequential models*, which identify ordered patterns, can be used to investigate whether specific sequences of temporally proximate discourse moves, such as talk or actions, explain variation in group outcomes (Kapur, 2011). *Cotemporal models*, which identify unordered but temporally proximate patterns, can be used to investigate whether discourse moves that co-occur within some period of time explain variation in group outcomes (Siebert-Evenstone et al., 2017).

Prior work has shown that both cotemporal models (Csanadi, Eagan, Shaffer, Kollar, & Fischer, 2019) and sequential models (Kapur, 2011) have advantages over atemporal models such as coding and counting; however, no empirical comparisons have been made between cotemporal models and sequential models in collaborative contexts. Thus, it is unclear whether accounting for temporal proximity, versus both temporal proximity and sequence, is a more effective approach for modeling collaborative processes. Such a comparison will help inform researchers about the conditions under which one class of models outperforms another, potentially impacting assessments of complex thinking and performance. But more importantly, comparing techniques that model temporality in terms of local sequence versus local cotemporality can provide insight in to the nature of temporality itself in CSCL contexts.

In this paper, we present an initial attempt to address this issue by comparing two specific modeling approaches: Epistemic Network Analysis (Shaffer, Collier, & Ruis, 2016), which models events cotemporally, and Sequential Pattern Mining (Srikant & Agrawal, 1996), which models events sequentially. We use both models to analyze data collected from air defense warfare (ADW) teams participating in computer-simulated training scenarios. We compare the efficacy of these models at finding differences in group performance in terms of (a) explanatory power, (b) efficiency, and (c) interpretability.

### Theory

Researchers in CSCL argue that collaboration has an important temporal dimension. For example, Kozlowski & Illgen (2006) argue that repeated interactions between individuals create behavioral, cognitive, or motivational states that influence future interactions. Similarly, Clark (1996) argues that as collaborative activities unfold in time, information is added to the *common ground*, the set of shared knowledge and experiences that exist between people when they interact, which in turn influences subsequent actions and interpretations (Dillenbourg, 1999).

Models of collaboration that do not account for temporality thus omit crucial information, limiting their validity (Kapur, 2011). One prevalent response to this critique has been to focus on *sequences* of discourse moves. Sequence is potentially important because, as Reimann (2009) argues, “human learning is inherently cumulative, [and] the sequence in which experiences are encountered affects how one learns and what one learns.” Moreover, in some collaborative settings, discourse moves should be carried out in a particular order. For example, Hutchins’ (1995) study of quartermasters in the U.S. Navy showed that navigation teams needed to follow a particular sequence of actions in order to accurately track the position of their ship. *Sequential models*, such as Sequential

Pattern Mining (SPM), have been used to identify sequences of discourse moves that groups make during collaborative processes in a variety of contexts (Perera et al., 2009).

There is, however, another aspect of the temporality of collaborative processes: *temporal proximity*. Events at any point in time are influenced by prior actions. However, the influence of prior activity does not always span the entire history of group interaction. Halpin and von Davier (2017) argue that the actions of one part of a group make actions by other parts more or less likely in the near future. For example, when one group member asks a question, others are likely to respond soon after. Suthers & Desiato (2012) argue that actions and interactions are interpreted with respect to the *recent temporal context*, or the immediately preceding events. Discourse moves within the recent temporal context may influence one another, directly reference one another, or build upon one another. Thus, while collaborative processes are composed of complex interactions among individuals, the most *relevant* interactions are bounded by temporal proximity.

Many SPM algorithms allow researchers to account for recent temporal context using sliding windows. Such algorithms add the constraint that the identified sequences must occur within a given window of events. However, an alternative approach is to model temporality based on co-occurrence rather than sequence, which focuses on temporal proximity irrespective of order.

In *cotemporal* models, two discourse moves are meaningfully connected if they co-occur within the same recent temporal context (Shaffer, 2017). For example, Epistemic Network Analysis (ENA) can be used to identify the connections groups make during collaboration (Shaffer et al., 2016). ENA identifies these connections by measuring how often particular discourse moves co-occur within the recent temporal context, operationalized as a sliding window that moves through each event—for example, problem step or turn of talk—in the data. ENA represents connections between discourse moves using undirected network models, meaning that connections between moves A and B in the network could mean that A followed B or that B followed A. In this way, ENA is *sensitive* to the order of events in the data—changing the order of events changes which events are present in a given window, and thus changes the results of the model—but the order in which discourse moves occur within any window is not represented in the model. ENA has been used to study CSCL and collaborative problem solving processes in many domains (e.g., Arastoopour, Shaffer, Swiecki, Ruis, & Chesler, 2016; Sullivan et al., 2018).

While it may seem counterintuitive to ignore the local order of discourse moves—after all, we perceive human actions as unfolding linearly in time—there are clearly contexts in which the specific order of these moves is less important than their cotemporality. For example, in complex and ill-formed problem solving, groups might consider issues A, B, and C at one point in the problem solving process; however, it may make little difference whether in that brief span of time they talk about A then B then C, or C then B then A—or any of the possible orderings of those issues. This approach has potential advantages over sequence models. For example, the results of SPM are lists of ordered patterns. A typical analysis identifies frequent patterns using SPM, then clusters those patterns using similarity metrics. Finally, researchers interpret these clusters in terms of the patterns within them (Jovanović, Gašević, Dawson, Pardo, & Mirriahi, 2017). In contrast, ENA produces network models for each unit of analysis, and provides an integrated visualization that helps to interpret the dimensions along which groups of networks differ. Because visualizations can reduce cognitive load in making inferences (Norman, 1993), cotemporal models with integrated visualizations may have interpretive advantages. Moreover, the number of possible permutations of discourse moves rises far more rapidly than their possible pairwise combinations as the number of significant types of moves increases. So in a case where specific ordering is not relevant, sequential models, which distribute sample variance across many variables, may have less explanatory power and may be overfit unless the analyzed dataset is large.

In this paper, we present an attempt to compare cotemporal and sequential approaches to modeling collaboration: ENA, which models events cotemporally, and SPM, which models events sequentially. We use both models to analyze data collected from air defense warfare teams (ADW) as they participated in computer-simulated training scenarios. In these scenarios, ADW teams detect, identify, assess, and take action toward nearby radar contacts. In theory, this process should follow a specific sequence for each contact; however, the demands of the task may lead to deviations from the sequence. Thus, the collaborative problem is neither so well-formed that only specific sequences are of interest nor so ill-formed that sequences are likely to be of little interest. We use this data to compare two hierarchical linear models—one using predictors derived from ENA and one using predictors derived from SPM—to assess whether ENA or SPM provides a more effective model of group performance in terms of (a) explanatory power, (b) efficiency, and (c) interpretability.

## Methods

As part of the Tactical Decision Making Under Stress project, 16 teams composed of six members each participated in four simulated training scenarios to test the impact of a decision-support system on team performance (Johnston, Poirier, & Smith-Jentsch, 1998). During the scenarios, teams performed the *detect-to-*



*engage sequence*. A *watch-station* provided basic information about identification and behavior of ships and aircraft in the vicinity (referred to as *tracks*). Teams needed to detect and identify multiple tracks, often simultaneously, assess whether they were threats, and decide how to respond, although the full sequence of actions did not apply to every track. (For example, non-hostile tracks did not require a response.) Teams in the control condition ( $n = 8$ ) had access to standard watch-stations. Teams in the experimental condition ( $n = 8$ ) had access to watch-stations enhanced with information about the tactical situation. Each team participated in the same four 30-minute scenarios; scenario order was counterbalanced using a Latin square.

We analyzed two data sources for each *team-scenario* (that is, each team in each scenario): (1) a transcript of team communications and (2) a teamwork behavior score. Transcripts were segmented into 12,027 turns of talk. Teamwork behavior was assessed in a prior study using the Air Defense Warfare Team Observation Measure (ATOM) (Johnston, Smith-Jentsch, & Cannon-Bowers, 1997), which summarizes four dimensions of teamwork performance—supporting behavior, leadership, information exchange, and communication—into an overall score from 1 (worst) to 55 (best).

## Coding

We analyzed the transcripts using the codes in Table 1, which were developed using a grounded analysis informed by both the existing ADW literature (e.g., Paris et al., 2000) and prior qualitative analyses conducted on similar data (e.g., Morrison et al., 1996). To code the data, we developed automated classifiers for each of the codes in Table 1 using the *ncodeR* package for the statistical programming language R (Marquart, Swiecki, Eagan, & Shaffer, 2018). We assessed concept validity by requiring that two human raters achieve acceptable values of Cohen’s kappa ( $\kappa > 0.65$ ) with statistically significant values of Shaffer’s rho ( $\rho < 0.05$ ), and we assessed reliability by requiring that both human raters independently achieve acceptable values of kappa and rho compared to the automated classifier (1). For each code, all pairwise combinations of raters (humans and automated classifier) achieved  $\kappa > 0.80$  and  $\rho(0.65) < 0.05$ .

**Table 1: Qualitative codes, definitions, and examples**

| Code                | Definition   | Example  |
|---------------------|--|--|
| Detection           | Talk about radar detection of a track or the identification of a track, (e.g., vessel type).                     | IR/EW NEW BEARING, BEARING 078 APQ120 CORRELATES TRACK 7036 POSSIBLE F-4                       |
| Track Behavior      | Talk about kinematic data about a track or a track’s location  | AIR/IDS TRACK NUMBER 7021 DROP IN ALTITUDE TO 18 THOUSAND FEET                                 |
| Assessment          | Talk about whether a track is friendly or hostile, the threat level of a track, or indicating tracks of interest | TRACKS OF INTEREST 7013 LEVEL 5 7037 LEVEL 5 7007 LEVEL 4 TRACK 7020 LEVEL 5 AND 7036 LEVEL 5. |
| Status Updates      | Talk about procedural information, e.g., track responses, or talk about tactical actions taken by the team       | TAO ID, STILL NO RESPONSE FROM TRACK 37, POSSIBLE PUMA HELO.                                   |
| Seeking Information | Asking questions regarding track behavior, identification, or status.  | TAO CO, WE’VE UPGRADED THEM TO LEVEL 7 RIGHT?  |
| Recommendation      | Recommending or requesting tactical actions  | AIR/TIC RECOMMEND LEVEL THREE ON TRACK 7016 7022   |
| Deterrent Orders    | Giving orders meant to warn or deter tracks.   | TIC AIR, CONDUCT LEVEL 2 WARNING ON 7037   |
| Defensive Orders    | Giving orders to prepare defenses or engage hostile tracks   | TAO/CO COVER 7016 WITH BIRDS   |

## Epistemic Network Analysis

To conduct a cotemporal analysis, we used the *rENA* package for the statistical programming language R (Marquart, Swiecki, Collier, et al., 2018). ENA uses a sliding window to construct a network model for each turn of talk in the data. Connections in the network are defined as the co-occurrence between codes in the current turn of talk and codes within the recent temporal context, which we defined as each line plus the four previous lines

based on our qualitative analysis of the data (a window size of 5 turns of talk). The resulting networks are aggregated for all turns of talk for each unit of analysis (team-scenario), such that each team-scenario is represented by a vector whose elements are the number of co-occurrences between each pair of codes for that team-scenario. ENA normalizes the matrix of co-occurrence vectors to account for variation in the amount of talk between teams and performs a dimensional reduction on the matrix via singular value decomposition.

Networks were visualized using two coordinated representations: (1) an *ENA score*, which represents the location of a team-scenario's network in the space (or *ENA space*) created by the dimensional reduction, and (2) a weighted network graph in which the nodes correspond to codes, and the edges are proportional to the relative frequency of connection between two codes. The positions of the network graph nodes are fixed across networks, and those positions are determined by an optimization algorithm that minimizes the difference between the ENA scores and their corresponding network centroids. Thus, ENA scores toward the extremes of a dimension have network graphs with strong connections between nodes located on the extremes. As a result, dimensions in this ENA space distinguish team-scenarios in terms of cotemporality between codes whose nodes are located at the extremes.

## Sequential Pattern Mining

To conduct a sequential analysis, we used the *TraMineR* package for the statistical programming language R (Gabadinho, Ritschard, Müller, & Studer, 2011). SPM identifies frequent patterns using a *support threshold*, where the support for a given pattern is the percentage of sequences that contain at least one instance of the pattern. We defined a sequence as the ordered list of codes across all turns of talk for a given team-scenario. In our data, codes can co-occur within a single turn of talk, so the SPM algorithm treats these codes as occurring simultaneously by defining *event sequences* which may contain patterns of un-ordered as well as ordered events (Ritschard, Bürgin, & Studer, 2013). We used a support threshold of 0.75 to limit the total number of patterns returned by the algorithm, which eases interpretation, and a window size of five turns of talk to match our ENA model (2). To interpret the SPM results, we counted the frequency of each high-support pattern for each team-scenario and applied Principal Components Analysis (PCA) to this data, resulting in a *PCA score* for each team-scenario—that is, the position of each team-scenario on the PCA dimensions (3). To interpret the PCA results, we used the dimension loadings, which show how much each variable—in this case, each frequent pattern—contributes to each dimension.

## Model comparison

We compared cotemporal and sequential models by constructing two Hierarchical Linear Models (HLMs). HLM is a regression technique for data with a nested-structure: in this case, team-scenarios (level-one) were nested into teams (level-two). In both HLMs, teams were random effects, the team-scenario ATOM score was the outcome variable at level-one, and Scenario was a control variable at level-two. For the cotemporal HLM (CT-HLM), ENA scores were explanatory variables at level-one. For the sequential HLM (S-HLM), PCA scores were explanatory variables at level-one.

We assessed model fit using two estimates of the total variance explained for HLMs: one from Snijders and Bosker (2012, TVE1) and the other from LaHuis and colleagues (2014, TVE2) (4). We assessed the efficiency of the models (model fit adjusted for number of parameters) using the Akaike information criterion corrected for small sample sizes. Following Burnham and Anderson (2004), we used a minimum  $AIC_c$  difference of 4 to indicate that the models were distinguishable.

## Results

### Epistemic Network Analysis

The first six ENA dimensions accounted for more variance in the data than any original variable. To reduce the chance of overfitting, we only used ENA scores on the first two dimensions in the CT-HLM. These two dimensions accounted for the highest proportion of the total variance: 51%. Figure 1 shows the average network across all team-scenarios and the ENA scores for each team-scenario in this space.

On the left side of the space are connections to Defensive Orders, Deterrent Orders, Status Updates, and Recommendations, all of which relate to actions taken by teams toward tracks. On the right side of the space are connections to Seeking Information. This suggests that the first dimension distinguishes team-scenarios in terms of whether they focused on *Tactical Actions* versus *Seeking Information*. Toward the top of the space are connections to Detection and Track Behavior; toward the bottom are connections to the remaining codes. Detection and Track Behavior relate to passing information about tracks, while the remaining codes relate to using

information about tracks. This suggests that the second dimension distinguishes team-scenarios in terms of whether they focused on *Track Information* versus *Track Processing*.

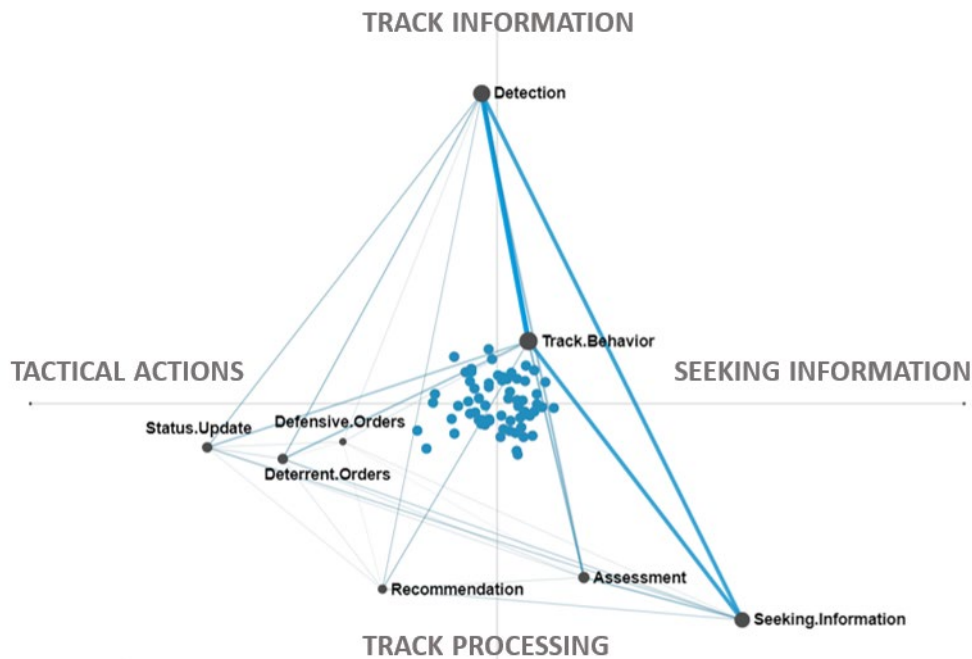


Figure 1. Average ENA network across all team-scenarios and ENA scores for each team-scenario.

### Sequential Pattern Mining

Our SPM analysis returned 165 patterns. PCA on counts of these patterns returned 23 dimensions that accounted for more variance than any original variable. To maintain consistency with the ENA analysis, we only used PCA scores on the first three dimensions, which accounted for 53% of the variance. As there are 165 original variables, interpretation requires considering 165 loadings for each dimension. However, it is common to use only the loadings with high magnitudes for interpretation. We interpreted each dimension by considering commonalities between the ten patterns that loaded at either extreme. To conserve space, we show only the three patterns at each extreme for each dimension in Table 2.

Table 2: Patterns with extreme loadings on first three principal components

| PC1   | PC2   | PC3   |
|---|---|---|
| <b>Patterns with Extreme Negative Loadings</b>    |   |   |
| (Track Behavior)-(Detection)-<br>(Track Behavior) | (Assessment)-(Seeking<br>Information)                               | (Track Behavior, Assessment)-(Seeking<br>Information)             |
| (Track Behavior)-(Detection,<br>Track Behavior)   | (Seeking Information)   | (Deterrent Orders)  |
| (Track Behavior)-(Track<br>Behavior)-(Detection)  | (Seeking Information)-<br>(Seeking Information)-(Track<br>Behavior) | (Status Update)-(Assessment)                                      |
| <b>Patterns with Extreme Positive Loadings</b>    |   |   |
| (Defensive Orders)                                | (Detection, Track Behavior)-<br>(Deterrent Orders)                  | (Seeking Information)-(Detection)-<br>(Detection, Track Behavior) |
| (Status Update)-<br>(Recommendation)              | (Detection)-(Deterrent Orders)                                      | (Seeking Information)-(Detection)-<br>(Detection)                 |
| (Status Update)-(Deterrent<br>Orders)             | (Status Update)-(Detection)   | (Detection)-(Detection)   |

The extreme negative side of the first dimension includes patterns involving Track Behavior and Detection. The extreme positive side includes patterns involving Defensive Orders, Deterrent Orders, Status Updates, and Recommendations. This suggests that the first dimension distinguishes team-scenarios in terms of whether they focused on *Track Information* versus *Tactical Actions*. The extreme negative side of the second dimension includes many patterns involving Seeking Information, while the extreme positive side includes patterns involving Status Updates or ending with Deterrent Orders. This suggests that the second dimension distinguishes team-scenarios in terms of whether they focused on *Seeking Information* versus *Deterring Tracks*. Finally, the extreme negative side of the third dimension includes many patterns involving Assessment. The extreme positive side includes patterns involving Seeking Information, Detection, and Track Behavior. This suggests that the third dimension distinguishes team-scenarios in terms of whether they focused on *Assessing Tracks* versus *Exchanging Information*.

## Model comparison

We compared the CT-HLM (ENA scores as explanatory variables) to the S-HLM (PCA scores as explanatory variables). Coefficients, standard errors (with corresponding *p*-values), as well as measures of model fit and efficiency for both models are shown in Table 3.

Table 3: Cotemporal HLM and sequential HLM, including parameter estimates and model fit

| Model  | Intercept               | ENA                     |                 | PCA            |                        |                 | Scenarios      |                 |                | Model Fit |       |                           |
|--------|-------------------------|-------------------------|-----------------|----------------|------------------------|-----------------|----------------|-----------------|----------------|-----------|-------|---------------------------|
|        |                         | 1                       | 2               | 1              | 2                      | 3               | B              | C               | D              | TVE1      | TVE2  | AIC <sub>c</sub>          |
| CT-HLM | <b>36.55*</b><br>(1.06) | <b>-10.1*</b><br>(2.77) | -0.42<br>(3.49) |                |                        |                 | 2.78<br>(1.58) | -2.34<br>(1.57) | 2.84<br>(1.49) | 0.259     | 0.253 | <b>377.19<sup>†</sup></b> |
| S-HLM  | <b>36.32*</b><br>(1.11) |                         |                 | 0.12<br>(0.09) | <b>0.38*</b><br>(0.18) | -0.18<br>(0.14) | 2.67<br>(1.61) | -1.61<br>(1.66) | 3.1<br>(1.55)  | 0.238     | 0.237 | 381.72                    |

( ) indicates standard error; \* indicates  $p < 0.05$ ; † indicates significantly lower AIC<sub>c</sub> score ( $\Delta > 4$ )

For the CT-HLM (cotemporal), the variance estimates for the random effects were 0.67 at level two and 16.70 at level one. Of the explanatory variables, only the coefficient for scores on the first ENA dimension (ENA1) was significant. The negative coefficient (-10.1) indicates that teams with higher teamwork behavior ratings focused more on *tactical actions* toward tracks than *seeking information about tracks*.

For the S-HLM (sequential), the variance estimates for the random effects were 0.06 at level two and 17.79 at level one. Of the explanatory variables, only the coefficient for scores on the second PCA dimension (PC2) was significant. The positive coefficient (0.38) indicates that teams with higher teamwork behavior ratings focused more on *deterring tracks* than *seeking information about tracks*.

For both estimates of total variance explained, the CT-HLM performed better. Moreover, the CT-HLM had an AIC<sub>c</sub> score 4.53 points lower than the S-HLM. Thus, the two models are distinguishable: the cotemporal model performed better than the sequence model in explaining differences in team behavior scores efficiently.

## Discussion

Our results suggest that cotemporal models of collaborative processes constructed with ENA outperform sequential models constructed with SPM on several dimensions. The cotemporal models explained more of the difference between team-scenarios, and while differences in total variance explained were small, model selection for efficiency via AIC<sub>c</sub> showed that the cotemporal-HLM was distinguishable from, and preferable to, the sequential-HLM. A possible explanation is that our ENA model used 28 un-ordered pairs of codes to derive predictors for the regression analysis, while our SPM model used 165 frequent patterns. The dimensional reduction on the ENA variables needed only *two* dimensions to account for a large proportion of the total variance while SPM needed *three* dimensions to account for approximately the same proportion of the total variance. Thus, the HLM model with ENA predictors was more parsimonious and less likely to overfit the data.

Our results also suggest that cotemporal models can have interpretive advantages compared to sequence models. The ENA algorithm combined connection identification, dimensional reduction, and visualization into one technique; once we generated the model it was ready to interpret. SPM, on the other hand, required several steps of post processing. More importantly, however, interpretation using ENA is done via integrated visualizations—network graphs projected into a low-dimensional space. SPM does not include such

visualizations. Because visualizations can reduce cognitive load by replacing cognitive calculations with perceptual inferences, ENA has an interpretive advantage compared to SPM.

Together, these results suggest that in CSCL and collaborative problem solving more generally, the specific *local order* of discourse moves may be less important than their *local cotemporality*. In the context we examined, teams were expected to follow a specific sequence of steps regarding each track; however, teams had to manage multiple tracks at once and the full sequence of actions did not apply to every track: it mattered more that specific discourse moves occurred together than that they occurred in a specific order.

Our results have several important limitations. First, we examined only one particular context of collaboration. However, the general features of this collaborative task, which contained both well-formed and ill-formed components, suggests that our findings may generalize to similar conditions in CSCL. Our future work will continue to investigate cotemporal and sequence models in similar contexts, as well as contexts that are more well-formed and more ill-formed. Second, it is possible that dimensional reduction on the SPM results via other techniques, such as clustering or factor analysis, could yield sequence models that perform better and are easier to interpret. Our results suggest that a two-cluster solution would be required to compete with ENA in terms of model efficiency, but each cluster would then need to be interpreted using more than 80 patterns. Allowing more clusters could ease interpretation but sacrifice efficiency. Similarly, we could apply a factor analysis to the SPM results using a two-factor solution, but it is unclear whether the performance and interpretability of the resulting sequence model would improve. Thus, we hypothesize results similar to those reported here. Our future work will compare cotemporal models developed using ENA to sequential models developed using clustering or factor analysis on SPM results to test this hypothesis.

Despite these limitations, our comparisons suggest that in CSCL contexts, cotemporal models can outperform sequential models. More importantly however, our results suggest that in some CSCL contexts, local order appears to be less important than local cotemporality. These results have implications for research and assessment in CSCL—in contexts that share both well-formed and ill-formed features, cotemporal models are a potentially more effective means of assessing complex thinking and performance. In turn, these models may better inform pedagogy and learning in such contexts.

## Endnotes

- (1) Shaffer's rho is a Monte Carlo rejective statistic that quantifies Type I error for generalizing from a sample of data coded by two raters to the true rate of agreement.
- (2) There are no established guidelines for choosing a support threshold. We tested multiple support thresholds above and below 0.75, and the quantitative results of the HLM comparisons were similar in all cases.
- (3) Performing a cluster analysis on the high-support patterns is more commonly used to interpret the results of SPM; however, methods are not reliably available for event sequences. One implementation exists (Ritschard et al., 2018), but it has not been validated. PCA has a standard method of interpretation and is agnostic to the type of pattern.
- (4) Because these models are non-nested, we were unable to test for significant differences in model fit.

## References

- Arastoopour, G., Shaffer, D. W., Swiecki, Z., Ruis, A. R., & Chesler, N. C. (2016). Teaching and assessing engineering design thinking with virtual internships and epistemic network analysis. *International Journal of Engineering Education*, 32(3B), 1492–1501.
- Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods & Research*, 33(2), 261–304.
- Clark, H. H. (1996). *Using language*. Cambridge university press.
- Csanadi, A., Eagan, B., Shaffer, D. W., Kollar, I., & Fischer, F. (2019). When coding-and-counting is not enough: Using Epistemic Network Analysis (ENA) to analyze verbal data in CSCL research. *International Journal of Computer-Supported Collaborative Learning*.
- Dillenbourg, P. (1999). *Collaborative learning: Cognitive and computational approaches. advances in learning and instruction series*. ERIC.
- Gabadinho, A., Ritschard, G., Müller, N. S., & Studer, M. (2011). Analyzing and visualizing state sequences in R with TraMineR. *Journal of Statistical Software*, 40(4), 1–37.
- Halpin, P. F., & von Davier, A. A. (2017). Modeling Collaboration Using Point Process. In *Innovative Assessment of Collaboration* (pp. 233–247). Springer.
- Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT Press.
- Johnston, J. H., Poirier, J., & Smith-Jentsch, K. A. (1998). Decision making under stress: Creating a research methodology. In J. A. Cannon-Bowers & E. Salas (Eds.), *Making decisions under stress: Implications for individual and team training* (pp. 39–59). Washington, D.C.: American Psychological Association.

- Johnston, J. H., Smith-Jentsch, K. A., & Cannon-Bowers, J. A. (1997). Performance measurement tools for enhancing team decision making. In M. T. Brannick, E. Salas, & C. Prince (Eds.), *Team performance assessment and measurement: Theory, method, and application* (pp. 331–327). Hillsdale, NJ: Erlbaum.
- Jovanović, J., Gašević, D., Dawson, S., Pardo, A., & Mirriahi, N. (2017). Learning Analytics to unveil learning strategies in a flipped classroom. *The Internet and Higher Education*, 33, 74–85.
- Kapur, M. (2011). Temporality matters: Advancing a method for analyzing problem-solving processes in a computer-supported collaborative environment. *International Journal of Computer-Supported Collaborative Learning*, 6(1), 39–56.
- Kozlowski, S. W., & Ilgen, D. R. (2006). Enhancing the effectiveness of work groups and teams. *Psychological Science in the Public Interest*, 7(3), 77–124.
- LaHuis, D. M., Hartman, M. J., Hakoyama, S., & Clark, P. C. (2014). Explained variance measures for multilevel models. *Organizational Research Methods*, 17, 433–451.
- Marquart, C. L., Swiecki, Z., Collier, W., Eagan, B., Woodward, R., & Shaffer, D. W. (2018). rENA: Epistemic Network Analysis (Version 0.1.3).
- Marquart, C. L., Swiecki, Z., Eagan, B., & Shaffer, D. W. (2018). ncodeR (Version 0.1.2).
- McGrath, J. E., & Tschan, F. (2004). *Temporal matters in social psychology: Examining the role of time in the lives of groups and individuals*. Washington: American Psychological Association.
- Morrison, J. G., Kelly, R. T., Moore, R. A., & Hutchins, S. G. (1996). Tactical decision making under stress (TADMUS) decision support system, 13.
- Norman, D. A. (1993). *Things that make us smart: Defending human attributes in the age of the machine*. New York: Basic Books.
- Paris, C., Johnston, J. H., & Reeves, D. (2000). A Schema-Based Approach to Measuring Team Decision Making in a Navy Combat Information Center. In *The Human in Command* (pp. 263–278). Springer, Boston, MA.
- Perera, D., Kay, J., Koprinska, I., Yacef, K., & Zaïane, O. R. (2009). Clustering and sequential pattern mining of online collaborative learning data. *IEEE Transactions on Knowledge and Data Engineering*, 21(6), 759–772.
- Reimann, P. (2009). Time is precious: Variable- and event-centred approaches to process analysis in CSCL research. *International Journal of Computer-Supported Collaborative Learning*, 4(3), 239–257.
- Ritschard, G., Bürgin, R., & Studer, M. (2013). Exploratory mining of life event histories. In J. J. McArdle & G. Ritschard (Eds.), *Contemporary Issues in Exploratory Data Mining in the Behavioral Sciences* (pp. 221–253). New York: Routledge.
- Ritschard, G., Studer, M., Bürgin, R., Gabadinho, A., Fonta, P.-A., Müller, N., & Rousset, P. (2018). TraMineRextras: TraMineR extension (Version 0.43).
- Shaffer, D. W. (2017). *Quantitative ethnography*. Madison, WI: Cathcart Press.
- Shaffer, D. W., Collier, W., & Ruis, A. R. (2016). A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics*, 3(3), 9–45.
- Siebert-Evenstone, A., Arastoopour Irgens, G., Collier, W., Swiecki, Z., Ruis, A. R., & Williamson Shaffer, D. (2017). In Search of Conversational Grain Size: Modelling Semantic Structure Using Moving Stanza Windows. *Journal of Learning Analytics*, 4(3), 123–139.
- Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Thousand Oaks, CA: Sage.
- Srikant, R., & Agrawal, R. (1996). Mining sequential patterns: Generalizations and performance improvements. *Advances in Database Technology—EDBT*, 1–17.
- Sullivan, S. A., Warner-Hillard, C., Eagan, B. R., Thompson, R., Ruis, A. R., Haines, K., ... Jung, H. S. (2017). Using epistemic network analysis to identify targets for educational interventions in trauma team communication. *Surgery*, in press.
- Suthers, D. D., & Desiato, C. (2012). Exposing chat features through analysis of uptake between contributions. In *System Science (HICSS), 2012 45th Hawaii International Conference on* (pp. 3368–3377). IEEE.

## Acknowledgments

This work was funded in part by the National Science Foundation (DRL-1661036, DRL-1713110), the U.S. Army Research Laboratory (W911NF-18-2-0039), the Wisconsin Alumni Research Foundation, and the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison. The opinions, findings, and conclusions do not reflect the views of the funding agencies, cooperating institutions, or other individuals.

# Teacher Monitoring Routines: Understanding Pedagogical Judgments During Students' Collaborative Learning

Nadav Ehrenfeld and Ilana S. Horn (1)  
Nadav.Ehrenfeld@vanderbilt.edu, Ilana.Horn@vanderbilt.edu  
Peabody College, Vanderbilt University

**Abstract:** In this paper we conceptualize *teacher monitoring routines* – a consequential yet understudied phase of instruction during students' collaborative work. We examine 5 lessons of experienced and equity-oriented secondary math teachers using a comparative case study design and analysis of their interactional routines. Our goal is to understand how teachers monitor groupwork and how they decide (intentionally or not) when, where, and how to intervene in student talk. We identified various patterns that make clearer the micro-judgments monitoring routines entail for teachers' (a) ways of moving among students and conversational *initiation*, (b) conversational *entry*; (c) *focus* of the interaction; (d) when and how to *exit* the interaction; and (e) conversation participation pattern; with an eye toward teachers' goals and instructional contexts. We end by considering implications for research on teaching and professional education in support of collaborative learning.

## Purpose

As mathematics teachers shift toward what has been called “ambitious instruction” (Lampert et al., 2011), they frequently build on collaborative groupwork to promote students' sensemaking of mathematical ideas through interaction (Cohen & Lotan 2014; Horn, 2012; Smith & Stein, 2011). While teachers' judgments during groupwork involve managing tensions between intervention and student autonomy, these micro-judgments remain understudied and ill defined (Stein et al., 2008; Webb et al., 2009). Different scholars who study this phase of lessons suggest different roles for teachers, with some advocating that teachers primarily listen to conversations (Smith & Stein, 2011) or intervene only when a group is “hopelessly off task” (Cohen & Lotan 2014), while others suggest it is more productive for the teacher to be more active and probe students' thinking (Webb et al., 2009).

Our own interest in studying teacher monitoring routines did not arise as a theoretical gap to fill, but rather as an *in vivo* problem of practice we encountered as teacher educators. For the past several years, we have partnered with a professional development organization (PDO) to bridge formal professional development with classroom teaching through video-based coaching. As we have spent time in the teachers' classrooms and discussed their teaching with them, we have noticed two things: (1) There is a surprising variability across teachers' groupwork monitoring practices; and (2) The “shakiest” moments of lessons tend to occur in the monitoring phase. While tasks, launches, and instructional routines can be easily shared among teachers, monitoring is inherently interpretive and responsive, requiring more of the teachers' sensemaking in-the-moment. The uncertainties involved in this phase of the lesson, along with the imperative to support cognitive demand and productive engagement for all students, provoked our interest in identifying the sequence of judgments teachers make during this phase of their lessons. To this end, we offer a framework for teachers' *monitoring routines* as a sequence of moves, which we name *Initiation-Entry-Focus-Exit*, to better understand the choices teachers make. Building on examples from 5 secondary math teachers, this paper aims to probe and conceptualize teacher monitoring routines, and to make clearer the myriad micro-judgments (intentional or not) they entail for teachers.

## Theoretical perspectives

### Interactional routines

In their basic form, interactional routines are recurrent *moves* constituting the patterned ways conversations unfold within a certain social group, as within a classroom. However, while routines might have some recognizable structure, interactional routines are also *emergent*, as their details are contingent on the situation. In this way, interactional routines are both stable and performative, reflecting structure in their consistent aspects as well as the agency of those who use them (Feldman & Pentland, 2003; Horn & Little, 2010; Lavie et al., 2018).

Attuning to the tensions of structure and agency, our conceptualization of monitoring routines builds on two definitions. The first comes from Cazden (2001), who describes interactional routines as holding both *sequential* and *selectional* dimensions. Cazden compares these two dimensions to a Western restaurant menu, likening the sequential dimension to the set of categories (e.g. appetizer, entrée, dessert) that is culturally set and rarely varies, while the selectional dimension is the specific appetizer, entrée or dessert that a person orders. For

monitoring routines, the structural, sequential dimension is captured in our generic framework of initiation, entry, focus, and exit, coupled with students access to participation in the interaction. The agentic, selectional dimensions are then the particular choices teachers make as they initiate, enter, focus, and exit interactions, based on their interpretations and responses to the situation.

For that reason, we find it compatible to use Lavie and colleagues' (2018) definition of interactional routines as a *task-procedure pair*:

“a routine performed in a given task situation, by a given person, is the task as seen by the performer, together with the procedure she executed to perform the task.” (Lavie, Steiner & Sfard, 2018, p. 9).

For example, in the case of groupwork monitoring routines, one teacher might see the *initiation move* as a task in which it is the teacher's responsibility to check-in with the groups. Another teacher might interpret the same task with different reasoning — for instance, as a situation where teacher-initiated conversation might interrupt students' thinking and thus would only approach a group if students had a question. Either way, some relatively stable structure of local classroom culture would be constructed, and students would know whether they need to raise their hands or simply wait for the teacher to come over. As we mentioned before, such teacher sensemaking is necessarily situated in the particularities of their teaching situations, including the teachers' instructional goals and their institutional constraints.

### Situated view of teaching

Teaching is an irreducibly situated act, as teachers navigate complex relational and institutional terrain during instruction. By taking a situated view, we acknowledge teachers' instruction as co-constructed with the particularities of their teaching situations (Greeno, 1998; Horn & Kane, 2015; Lave & Wenger, 1991). As a consequence, teachers' monitoring practices are an interactional accomplishment between teachers and the particular students, content, and settings they are working with. Although routines may stay somewhat stable within individual teachers as they move between classrooms or groups, we also see teachers making micro-judgments about whom they direct their attention to, in what ways, and for how long. For this reason, while abstract and idealized notions of monitoring practices may serve as a useful heuristic guide to practice, teachers necessarily exercise *pedagogical judgment* as they enact any instructional practice in their teaching situations (Horn, in press).

Applying this lens to the teachers in our study, we are especially interested in how their pedagogical judgment shapes the ways they monitor student worktime. By highlighting teachers' micro-judgments as they monitor students' work, we claim that teachers do more than simply implement known models of groupwork; teachers are constantly assessing, interpreting, and adjusting as groupwork unfolds in their particular classroom contexts.

### Research question

Given our goal to probe and conceptualize teacher monitoring routines and considering our theoretical perspective, we ask: *What interactional routines emerge during the focal teachers' conversations with student groups during the monitoring phase of the lessons?* In the rest of the paper, we answer this question, with an eye towards the teacher micro-judgments that underlie these routines.

## Data and methods

### Research context and data

As we mentioned, this study comes from a research practice partnership between our university research team and a PDO. Our shared work centers on bridging formal professional development and classroom instruction, and we have worked to design a video-based formative feedback (VFF) coaching cycle to help elicit and engage secondary math teachers' pedagogical judgment. At the PDO, the secondary mathematics teachers have encountered different models of monitoring groupwork, including the aforementioned *5 Practices* (Smith & Stein, 2011) as well as Complex Instruction (Cohen & Lotan 2014; Horn, 2012). In addition, they have attended workshops on the TRUmath Framework (Schoenfeld, 2014), which highlights issues of student agency and access to rich mathematics. In sum, our participants have had atypically rich opportunities to learn about ambitious instruction in formal professional development.

For this analysis, we looked at 5 lessons (see Table 1) from experienced secondary teachers in the PDO (5 to 18 years of teaching experience, mean = 8 years), in a large urban district in the western U.S. They were



selected because they included segments of 20 minutes or longer of teachers' monitoring students' mathematical groupwork (20 to 45 min segments, mean = 28.5 min). For all 5 lessons, we have two cameras recording the class session. Camera 1, a tablet camera on a robot tripod, captured the whole class, with a focus on the teachers' movements through the classroom. These video records also have student audio tracks captured through 4 separate microphones placed at students' desks. Camera 2, a point of view camera, was mounted on the teacher's head, shoulder, or chest to approximate what they were seeing as they moved through the classroom and interacted with students. In addition to these recordings, our data include lesson artifacts, conversations with the focal teachers before and after instruction, and email exchanges about the activities.

Table 1: information about the focal teachers and the lessons

| Focal teacher      | Experience (years) | School   | Class                      | Date        | Groupwork (minutes) (%) |
|--------------------|--------------------|----------|----------------------------|-------------|-------------------------|
| Veronica Kennedy   | 8                  | Rees     | 8 <sup>th</sup> grade math | Feb 1 2018  | 33 of 82 (40%)          |
| Brad Miller        | 6                  | Noether  | Algebra 1                  | May 7 2018  | 20 of 58 (34%)          |
| Bridgette Campbell | 7                  | Johnson  | Algebra 1                  | Oct 05 2017 | 20 of 85 (23%)          |
| Lee Bellver        | 14                 | Falconer | Algebra 1                  | Mar 1 2018  | 26 of 100 (26%)         |
| Lizette McLoughlin | 5                  | Fermat   | AP Calculus                | Feb 2 2018  | 45 of 55 (82%)          |

## Data analysis

### Phase 1: Data reduction and re-representation

We edited the video to isolate the monitoring sequences of each lesson, reducing 380 min of classroom footage to 144 min. Then, using Final Cut Pro X, we edited the two camera perspectives together as a picture-in-picture to support coordinated analysis of the teachers' actions and a view of where they are located in the classroom. (See Figure 1.)

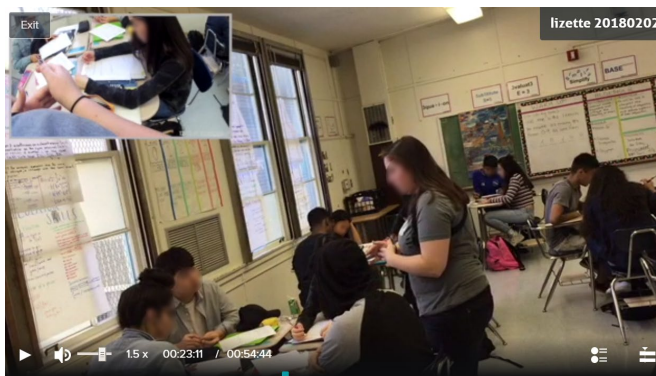


Figure 1. Two camera perspectives on Lizette's classroom.

### Phase 2: The sequential dimension of teachers' monitoring routines

Using the subset of video data from Phase 1, we used inductive coding (Strauss & Corbin, 1998) to identify the sequential dimension of the *monitoring routine*. We arrived at categories that explained all 5 cases, and they were 5 recurrent *moves* that described: (a) how teachers moved around the room and how conversations were initiated (*initiation*) (b) their opening moves of conversation (*entry*); (c) the main *focus* of the interaction; (d) when and how they *exit* the interaction; (e) whether the teacher interacts with students as a group or as individuals (see Table 2). This constituted the general task of the emergent monitoring routines.

### Phase 3: The selectional dimension of teachers' monitoring routines

Once we arrived at the basic sequence of monitoring routines, we could delve into the interactional details — the particular judgments teachers made in their teaching situations. Using an inductive comparative case study design (Strauss and Corbin, 1998) and methods of interaction analysis (Jordan and Henderson, 1995), we looked closely at the videos to explore patterns of how the 5 teachers took up each of the 5 moves identified in Phase 3. The coding team consisted of two PhD students (including the first author) and a secondary math teacher. During the coding process, the team met regularly, maintained an online “coding dilemmas” document and a codebook with definitions and examples for each code. As coding dilemmas were discussed and resolved among the coding team

and the Principle Investigator (the second author), the definitions and examples were refined. Peer Debriefing was done with the larger research team (2 additional PhD students, a post-doctoral fellow, and an assistant professor) to ensure the ecological validity of the coding scheme (Lincoln & Guba, 1985).

Once the codes were more or less agreed upon, the team used Vosaic software to code the videos. Python code (in Pycharm framework) was used to process the data and to create representations that account for the teachers' time spent with each student group as well as the sequence of their visits (Figures 2,4,5).

Table 2: description of the 4 basic moves and participation pattern

| Move                         | Description   | Codes   |
|------------------------------|---|---|
| <b>Initiation</b>            | How the teacher approaches the group?                   | <ul style="list-style-type: none"> <li>• <i>Student-initiated</i>: Student calls teacher over</li> <li>• <i>Teacher-initiated</i>: The teacher approaches the group of their own accord</li> </ul>  |
| <b>Entry</b>                 | Teacher's first statement to the group                  | <ul style="list-style-type: none"> <li>• Teacher asks for a summary or "what Ss are up to" in terms of math or participation norms</li> <li>• Teacher answers students' question</li> <li>• Teacher specifically asks about results or pacing</li> <li>• Teacher redirects Ss interaction to a new topic</li> </ul>   |
| <b>Focus</b>                 | The substance of the interaction                        | <ul style="list-style-type: none"> <li>• Teacher discusses <i>mathematical ideas</i> <ul style="list-style-type: none"> <li>○ Probing (how? why?)</li> <li>○ Discussing results (what?)</li> <li>○ Hinting (pointing resources or evaluating students' ideas)</li> </ul> </li> <li>• Teacher discusses <i>group dynamics</i></li> <li>• Teacher discusses <i>task directions</i></li> <li>• Technical issues or brief comment</li> <li>• <i>Any combination</i> of the above</li> </ul> |
| <b>Exit</b>                  | The last directive the teacher gives before moving away | <ul style="list-style-type: none"> <li>• <i>Closed</i> directive of "next steps"</li> <li>• <i>Open</i> directive of "things to think about"</li> </ul>   |
| <b>Participation pattern</b> | Whom does the teacher interact with?                    | <ul style="list-style-type: none"> <li>• Teacher discussing with <i>whole group</i></li> <li>• Teacher discussing with <i>individuals or subgroup</i></li> </ul>  |

## Findings

As teachers monitor groupwork, we noted five distinct types of moves, generally following the sequence of *initiation*, *entry*, *focus*, and *exit*, as well as an overarching *participation pattern* summarized in Table 2. In the following section, we discuss each of these, using two contrasting cases from our data set to illuminate and span a space of possibilities by which these moves can be taken up.

We offer three caveats about the scope of this analysis. First, our intention is not to be exhaustive: We do not claim that these contrasts capture the full breadth of how teachers might use the moves to monitor groupwork. Second, we do not intend to claim that these routines capture an essence of any individual teacher. In fact, we have examples in our data set of the same teachers leading different lessons where their monitoring routines look quite different. Finally, we do not seek global, normative statements about monitoring routines. Returning to our situative perspective, we see the monitoring routine's structure as capturing a wide range of instruction across a variety of teaching situations. Instead of highlighting "best practices," we identify these phases to help teachers reflect on and deliberate on the range of micro-judgments they can make during this crucial phase of the lesson.

## Initiation

We used three codes for the initiation phase of each group interaction describing who initiated the teachers' involvement: *student(s)*, *teacher*, or *unclear*. (This third code was not used much.) Within each monitoring case, we saw patterns of *mostly student initiation*, *mostly teacher initiation*, or *mixed initiation*. These top-level patterns capture the general tendency of different teachers to intervene in student groupwork as well as the degree of student autonomy in the class.

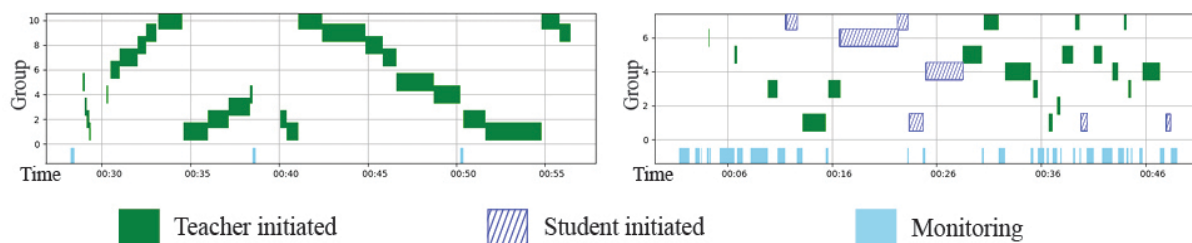
To illustrate how initiation patterns vary, we contrast Veronica and Lizette (Figures 2a and 2b). Because these are nonstandard representations, we offer this guide to understanding what they mean: the x-axis represents

time, the y-axis represents each unique group in the class, with -1 representing moments when the teacher did not directly interact with a group. Thus, the rectangles represent discrete interactions, and their lengths represent the amount of time the teacher spent with a particular group. Green rectangles signal teacher-initiated interactions, while blue rectangles signal student-initiated ones. Immediately we see that Veronica's top-level pattern was *mostly teacher initiated*, while Lizette's was *mixed initiation*.

Using these representations, we note a few differences between Veronica and Lizette's initiation routines. First, while Lizette spent 31% of her time *quietly monitoring* (-1 on the y-axis, light blue rectangles), Veronica only spent 2% of hers in the same way. Turning to the underlying pedagogical judgment, we surmise that this may be due to the age difference between the students and, relatedly, their degree of autonomy: Veronica taught 8<sup>th</sup> grade math, while Lizette taught older students AP Calculus. From our conversations with the teachers, we also see these differences as arising from their conceptions of their role as a teacher during groupwork. For example, when we discussed Veronica's monitoring pattern with her, she shared that she did not think about standing back and watching.

Second, Veronica walked around the room in a fixed pattern from group to group, represented by the sinusoidal curve, whereas Lizette constantly changed the order that she approached student groups. Turning to the underlying pedagogical judgments, we can interpret Veronica's fixed pattern as informed by her judgment that she needs to distribute her attention evenly and help keep the students on task, whereas Lizette's irregular pattern might signal her responsiveness to students' initiation, which may be less predictable.

To reiterate, we do not seek to evaluate these moves as good or bad in an absolute sense. However, we are curious about the consequence of teachers' moves for students' opportunities for mathematical engagement. As we listened to the student audio tracks during Veronica's lesson, for instance, we noted several times that, as a consequence of her fixed pattern of teacher initiation, students' vibrant mathematical talk got interrupted by Veronica's intervention. At the same time, the fixed pattern results in a relatively equal time allocation among the groups, which, as the diagrams show, was not true for Lizette's students. For example, Lizette conversed with Group 2 briefly exactly once and conversed with Group 6 only twice, with the second time lasting much longer. Of course, equal time allocation is not necessarily equitable or even productive for learning. A group that is working well might not need the teacher intervention. But all this points back to the centrality of pedagogical judgment during monitoring: What 'working well' means, for whom, and when, is a matter of interpretation. However, knowing that *not to intervene* is a sound possibility — and one that hadn't occurred to Veronica — might help support the complexity of this (intentional or not) micro-judgment.



Figures 2a & 2b. Veronica's (left) & Lizette's (right) initiation routines.

## Entry and focus

Entry and focus capture the heart of the teachers' interactions with student groups. First, the four codes that captured teachers' entry into the conversation answered the question: *how does the teacher verbally begin the interaction with the students?* Our codes were: (1) *listening/asking for a summary*; (2) *asking about results*; (3) *answering student questions*; and (4) *redirecting interaction*. The focal teachers typically started their entry by asking where the group stands (codes 1 and 2). Code 1, listening or asking for a summary, can center on either in terms of their mathematical thinking ("so tell me about what you're doing") or participation patterns ("are they helping you out?").

Next, to capture the focus of the conversations, we asked, *what is the nature of the sensemaking throughout the conversation?* We had six codes for the focus, 3 of which involved math talk and 3 of which involved not-math talk. The three math codes were: *probing students' explanation*, *results-centered talk*, and *teacher hints/scaffolds*. The three not-math codes were: *norms of participation*, *assignment centered*, and *technical issues/brief asides*. Of our four moves in the monitoring routine, *focus* was the only one that was often double coded, as teachers commonly focused on more than one thing over the course of an interaction.

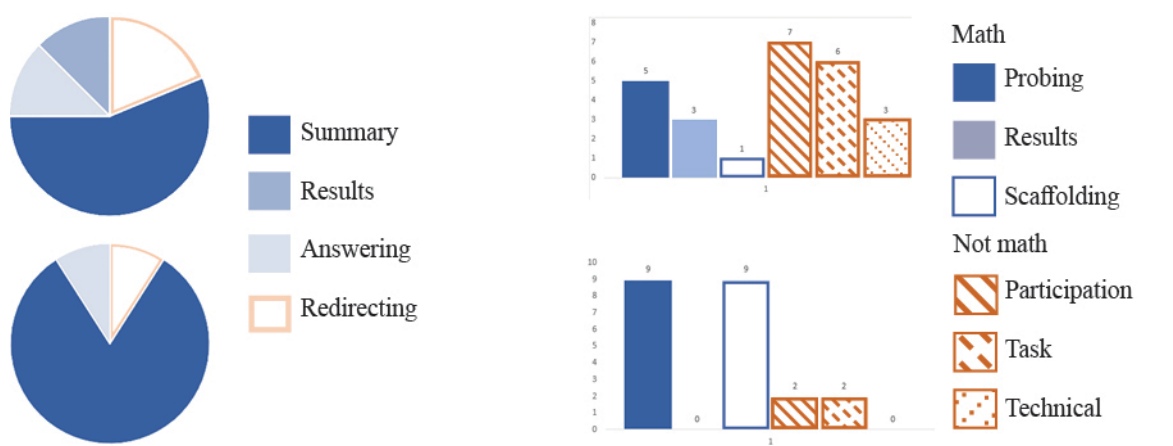
To illustrate how entry and focus can vary, we contrast Bridgette and Brad. Bridgette (Figure 3a) was the only teacher who spoke about participation more than she did about math. In comparison, Brad's entry and

focus (Figure 3b), was more typical, since most of his focus was mathematical, with a constant tension between probing students thinking and scaffolding it.

Like Lizette, Bridgette quietly monitored her classroom more than the other teachers in our sample. However, unlike Lizette (see Figure 2b), Bridgette seldom initiated interactions during the monitoring phase (only 3 out of 16 interactions were teacher-initiated). Bridgette also assigned group roles to students and was very consistent about redirecting student questions back to their peers. In sum, at a top-level description, Bridgette’s entry and focus aimed for *equitable participation*.

Brad, on the other hand, like most of our focal teachers, primarily spoke about math with student groups. Looking into the types of math talk, we see a tension between *probing students’ thinking* and *hints/scaffolds*. As we mentioned above, Brad’s conversations generally started with an entry move (move 2) of asking for a summary of what they have done, followed by a focus (move 3) on *probing* students’ questions and thinking, concluding with a strong *hint/scaffold* by either evaluating their thinking or directing students to a next step to continue with. In sum, at a top-level description, Brad’s entry and focus aimed for *guided work completion*.

By contrasting Bridgette’s equitable participation approach with Brad’s guided work completion approach, we again abstain from evaluation of their actions. Instead, we emphasize the tensions they were navigating in making the micro-judgments that inform their practice. For instance, Bridgette’s emphasis on equitable participation pressed on student involvement and centered their mathematical thinking. However, in the class session we observed, not all groups completed their task for the day, a valued outcome for many teachers. In contrast, Brad’s students, guided by his responsive hints and scaffolds, mostly completed their work, although its production may not have been as equally distributed across students. How and how much to balance probing and scaffolding, and for whom and when, again is ultimately a matter of teachers’ pedagogical judgment.



Figures 3a & 3b. Bridgette’s (upper charts) & Brad’s (lower charts) entry and focus routines.

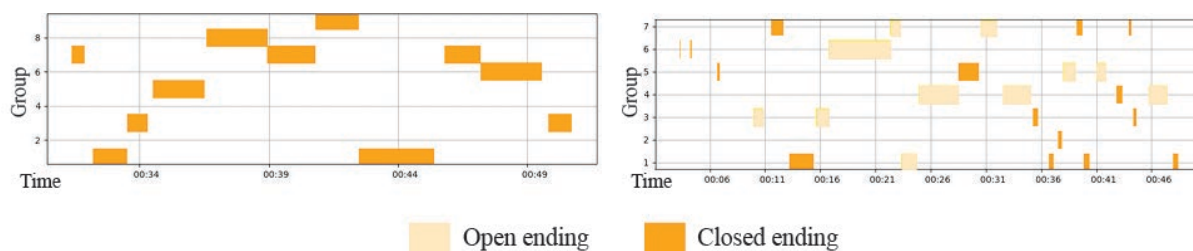
## Exit

To capture the exit moves, we asked, *how does the teacher end the interaction with students?* with an eye toward how varying exits set students up for continued conversation. For this reason, we distinguished between *open* exits that point toward further exploration, and *closed* exits, which involved clearer directives. For example, an open ending might offer an issue to think about, while a closed exit would tell a next step.

To illustrate exit moves, we contrast Brad and Lizette. Using diagrams constructed similar to those illustrating the initiation move, Figures 4a and 4b show the teachers’ movement around their classroom, the time spent with each group, with yellow rectangles signaling open exits and orange rectangles signaling closed exits. Here, we see that Brad’s exit routines (Figure 4a) provide us with important insights regarding his overall monitoring routine. On the one hand, Brad elicited students thinking and had rich mathematical conversations with the groups (which we omit here for lack of space). On the other hand, all of Brad’s 11 group interactions had a closed ending (e.g., “so the vertex should come in the middle”). In contrast to Brad, Lizette (Figure 4b) had mostly open exit moves, which included less directive hinting, such as pointing to a resource (e.g. “anything we’ve talked about last week that might help here?”), generally leaving the group with something to think about rather than something to do.

Once again, the teachers’ understanding of their role during monitoring shapes these moves. The distinction between open and closed endings are critical here, because it might have important implications for how students’ conversation will unfold after the teacher walks away. In Lavie and colleagues’ (2018) terms, if

teachers interpret their task as supporting students' collaboration in their absence, they should privilege open exits. If teachers interpret their task as wrapping up the conversation, ensuring work completion, or supporting student pacing, then they should privilege closed endings.



Figures 4a & 4b. Brad's (left) and Lizette's (right) exit routines.

## Participation pattern

As we were coding the moves of the monitoring routine, we noticed differences in the overarching participation pattern between the teacher and student groups. Specifically, whom the teacher interacts with during intervention and in what ways. To illustrate participation patterns, we turn to the contrasting cases of Bridgette and Lee, two outliers in our sample. Most of the other cases show teachers moving between interacting with all students in a student group or only with some. In many cases, teachers start a conversation with certain students and slowly (and seemingly unintentionally) exclude other students from the discussion through their gaze or body positioning.

Bridgette, as we have already described, was very intentional about equitable participation. Of our 5 focal teachers, Bridgette was the only one who consistently addressed all students in the group (16 of 16 interactions). Other teachers were either less intentional about this participation pattern or clearly facilitated a different one. Lee, for example, walked between groups and made sure to talk to every student; however, he addressed them one at a time. We see this as evidence that Lee's interpretation of his own task was less about promoting collaboration and more about supporting individual students' understanding of content.

## Discussion

Teachers' monitoring practices are consistent enough to be captured as emergent routines involving initiation, entry, focus, and exit, with general overarching participation patterns. We note two things about these routines. First, the different teacher monitoring routines described above have implications for sustaining students' mathematical engagement. Our claim is not that one mode of monitoring is always appropriate, but rather that monitoring groupwork entails complex interactions, some of which teachers may not be aware of. Second, the differences in monitoring routines seem to reflect the teachers' understandings of their role during groupwork. By examining and conceptualizing this understudied phase of instruction, we illuminate how different monitoring routines both shape the learning environment and are shaped by it.

Although we see monitoring practices as reflecting teachers' conceptions of their role, we also stress that they are not teachers' fixed "teaching style." For example, while Lizette was quietly monitoring 31% of the time in this class session and mainly used open-ended exits, she was teaching an AP Calculus class. If she, like Veronica, had been teaching a middle school mathematics class, she may not have had a chance to circulate quietly, and she may have exited conversations in a content oriented way to work toward task completion — or to quell the anxieties of the younger students. Other aspects of the teaching situation surely influence monitoring routines, such as the extent to which the tasks are groupworthy (Lotan, 2003) and thus invite open ended exploration, as well as the physical space of the classroom. For instance, some of our teachers used vertical whiteboards on the perimeter of the room for groupwork. In these classrooms, the sinusoidal pattern emerged fairly consistently as teachers circulated around the room's perimeter. (See Veronica [Figure 2a] and Lee [Figure 5b] for examples.) In contrast, teachers who are walking between desks could circulate quietly without officially being part of some group discussion.

While supports exist to aid teachers in building student-centered classrooms, this analysis highlights key points in this work, uncovering the range of possibilities for monitoring groupwork, and the complexity of the micro-judgments required in monitoring. As we better understand these routines, we can develop clearer ways to support productive equitable mathematical discourse during this crucial phase of lessons.

## Endnotes

(1) The authors contributed equally to this paper.

## References

- Cazden, C. (2001). *Classroom discourse: The language of teaching and learning*. Second Edition. Portsmouth, NH: Heinemann.
- Cohen, E. G., & Lotan, R. A. (2014). *Designing Groupwork: Strategies for the Heterogeneous Classroom*. Third Edition. Teachers College Press.
- Feldman, M. S., & Pentland, B. T. (2003). Reconceptualizing organizational routines as a source of flexibility and change. *Administrative Science Quarterly*, 48, 94–118.
- Greeno, J. G. (1998). The situativity of knowing, learning, and research. *American psychologist*, 53(1), 5.
- Horn, I. S. (2012). *Strength in Numbers: Collaborative Learning in Secondary Mathematics*. National Council of Teachers of Mathematics.
- Horn, I. S. (in press). Supporting the Development of Pedagogical Judgment: Connecting Instruction to Context. *International Handbook of Mathematics Teacher Education*, 2nd edition, Volume 3. Sense Publishers.
- Horn, I. S., & Kane, B. D. (2015). Opportunities for professional learning in mathematics teacher workgroup conversations: Relationships to instructional expertise. *Journal of the Learning Sciences*, 24(3), 373-418.
- Horn, I. S., & Little, J. W. (2010). Attending to problems of practice: Routines and resources for professional learning in teachers' workplace interactions. *American educational research journal*, 47(1), 181-217.
- Jordan, B., & Henderson, A. (1995). Interaction analysis: Foundations and practice. *The journal of the learning sciences*, 4(1), 39-103.
- Lampert, M., Boerst, T. A., & Graziani, F. (2011). Organizational Resources in the Service of School-Wide Ambitious Teaching Practice. *Teachers College Record*, 113(7), 1361-1400.
- Lave, J., Wenger, E., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge: Cambridge university press.
- Lavie, I., Steiner, A., & Sfard, A. (2018). Routines we live by: from ritual to exploration. *Educational Studies in Mathematics*, 1-24. <https://doi.org/10.1007/s10649-018-9817-4>
- Lincoln, Y. S., & Guba, E. G. (1985). Establishing trustworthiness. *Naturalistic inquiry*, 289-331.
- Lotan, R. A. (2003). Group-worthy tasks. *Educational Leadership*, 60(6), 72-75.
- Schoenfeld, A. H. (2014). What makes for powerful classrooms, and how can we support teachers in creating them? A story of research and practice, productively intertwined. *Educational researcher*, 43(8), 404-412.
- Smith, M. S., & Stein, M. K. (2011). *5 practices for orchestrating productive mathematics discussions*. Reston, National Council of Teachers of Mathematics.
- Stein, M. K., Engle, R. A., Smith, M. S., & Hughes, E. K. (2008). Orchestrating productive mathematical discussions: Five practices for helping teachers move beyond show and tell. *Mathematical thinking and learning*, 10(4), 313-340.
- Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Procedures and techniques for developing grounded theory*. Sage.
- Webb, N. M., Franke, M. L., De, T., Chan, A. G., Freund, D., Shein, P., & Melkonian, D. K. (2009). 'Explain to your partner': teachers' instructional practices and students' dialogue in small groups. *Cambridge Journal of Education*, 39(1), 49-70.

## Acknowledgments

We are thankful to the participating teachers, to Nicole Louie and Anna Sfard who gave helpful feedback on earlier stages of work, to Avital for her help with the graphic representations, and to the research team: Patricia Buenrostro, Grace Chen, Brette Garner, Mariah Harmon, Lara Jasien, Samantha Marshall, Elizabeth Metts, and Katherine Schneeberger McGugan. This work has been supported by a grant from the National Science Foundation.



# Impacts of Augmented Reality on Collaborative Physics Learning, Leadership, and Knowledge Imbalance

Iulian Radu, Harvard University Graduate School of Education, iulian\_radu@gse.harvard.edu  
Bertrand Schneider, Harvard University Graduate School of Education, bertrand\_schneider@g.harvard.edu

**Abstract:** Emerging technologies such as Augmented Reality (AR), have the potential to radically transform science education by making challenging concepts visible and accessible to novices. In this project, we designed a Hololens-based system in which participants learned about the physics involved in audio speakers (Radu & Schneider, 2019). We analyzed participant dyad where educational AR representations were present or not, focusing on the relationships between collaboration, learning, leadership, and knowledge imbalance. We find that, overall, AR representations improved time management, learning of structural concepts but reduced learning of physical concepts. The effects of leadership were mediated by the presence of AR: the presence of leaders in AR was linked to higher learning gains and better collaboration, whereas these effects were not present without AR. Finally, for groups with imbalanced knowledge, AR seemed to benefit participants in configurations where less-knowledgeable participants took the lead in discussions. These results indicate that AR can be beneficial for equalizing the effects of imbalanced collaboration. We discuss the implications of those results for the design of CSCL learning activities using augmented reality.

## Introduction

In this research we investigate the benefits and drawbacks of augmented reality for inquiry-based learning, specifically in relation to participant agency and leadership. We focus on a collaborative activity that allows pairs of students to explore concepts in electromagnetism. Electromagnetism is a topic that is often encountered in both maker spaces and traditional physics classrooms, and it is one of the most difficult topic to master because it combines understanding of physical objects (ex: magnets, wires) and abstract concepts (ex: magnetic field shapes, voltage, electricity) (Belcher and Bessette 2001; Ibáñez et al. 2014; Maloney et al. 2001).

We focus on how AR technology in this educational context relates to leadership and individual agency. A prevailing issue in classroom group collaborations is the effect of unequal participation and unequal knowledge among group members. Some people are naturally more dominant while others more passive, and this may be enhanced by the amount of each person's domain expertise (Salomon & Globerson, 1989). This imbalance can create negative effects whereby the more passive or less-knowledgeable students do not contribute as much as their peers, potentially leading to less effective problem solving, collaboration breakdowns, and "free rider" effects. Augmented reality environments have the potential to mediate the effects of participant leadership and knowledge imbalances. On one hand, AR environments can provide a plethora of holographic visualizations, which can be beneficial for participants by allowing passive or lower-knowledge participants to easily follow the communications of their peers; they allow participants to more easily ask for clarification or interrupt by pointing at representations; they allow more knowledgeable participants to explain by use of the available representations; and they can allow more equitable interaction with the experience, thus providing agency to participants who may naturally be more passive. On the other hand, AR environments require the use of specialized hardware, including the use of head-mounted devices which may cover the participants' eyes, thus making communication difficult by masking nonverbal cues such as facial expressions.

In this study we investigate these topics by studying dyads using head-mounted Microsoft Hololens AR devices, in conditions where educational AR representations are either present or not. We predict that the presence of AR information will balance collaboration because participants have shared access to information, thus one person will be less likely to dominate the experience; additionally, we expect the availability of shared representations to potentially increase the participation of more passive group members. For this reason, we predict that educational AR visualizations will have positive effects on collaborative processes. Finally, we observe how groups behave when more active participants (drivers) have high or low knowledge, and expect that groups with more knowledge imbalance will be negatively impacted in the condition where groups do not have access to educational AR information because a "free rider" effect (Salomon & Globerson, 1989).

## Related work

Augmented reality systems have been developed for a wide range of educational uses, such as for learning geometry (Radu et al. 2015), chemistry (Yu-Chien 2006), and history (Chang et al. 2015). Specifically for physics

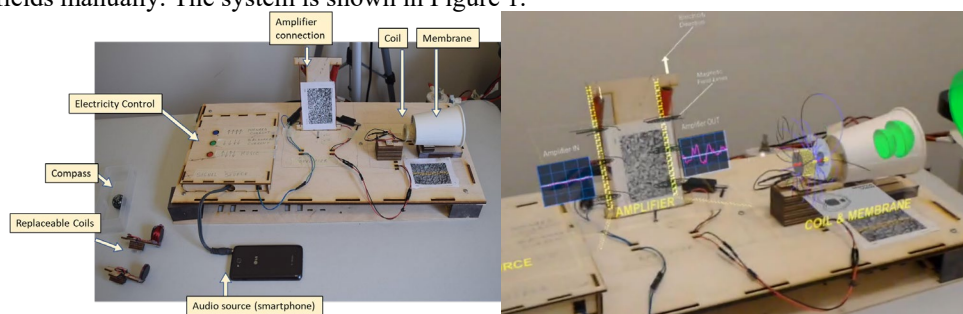
education, systems have been built for visualizing electricity (Belcher and Bessette 2001) and magnetism (Maloney et al. 2001). Augmented reality systems can support student learning of spatial structures, improve performance on physical tasks, and have lasting effects on participant memory (for a comparative media review see Radu, 2014). While the effects of AR experiences on learning have been studied, there is relatively less research focused on understanding the impact of AR on the dynamics of co-located collaboration. Previous research projects have focused on designing AR infrastructures that support social interactions (Billinghurst, Clark, and Lee 2015), for example by allowing a remote expert to inhabit a physical space, allowing multiple users to annotate physical environments, and allowing users to have different views of the same environment. Research has also found that student groups using augmented reality can benefit from increased collaboration because they are more motivated to engage with the experience (Phon, Ali, & Halim, 2014) and because the experiences simulate real-world professional collaborations (Dunleavy & Dede, 2014). Furthermore, it has been argued that AR experiences can balance leadership behaviors of participants, since group members can have access to shared visualizations and one person is less likely to control the group resources (Morrison et al., 2009).

We contribute to this research agenda by specifically studying how collaboration aspects of leadership and knowledge imbalance are impacted by the presence of educational AR, presented as holograms on physical artifacts. A prevailing issue in group collaborations is the effect of unequal participation among group members. Unequal participation is caused by multiple factors including unequal knowledge relevant to the activity, unequal ability to control the activity, or unequal personal interest and initiative. In situations where team members do not (or cannot) contribute equally, this typically results in lower learning gains (Chen 2006). Previous research indicates that when a resource is limited among team members, this encourages one person to dominate the interaction and creates imbalanced participation (Church, Hazlewood & Rogers 2006). In such contexts, participants may simply follow along, leading to decreased learning gains and poor collaboration (Salomon and Globerson 1989). Shared interfaces such as tangible objects have the potential to balance participation as each user has access to the learning content (Church, Hazlewood & Rogers 2006), especially when such interfaces allow participants to have shared control and awareness of information available to the group (Yuill and Rogers 2012).

## Augmented reality system and research questions

In this section, we describe how prior work informed the design of our system, the study we conducted, our research questions and the methodology we used to answer them.

We have designed a tangible model of a sound-producing speaker, augmented with interactive educational holograms visible to two users through the Microsoft HoloLens augmented reality headset. This educational system mixes physical content (physical wires, magnets, magnetic forces, movement) and virtual content (visualizations of electricity, magnetic fields, force directions). The participants can interact with the system by playing music or constant electric current in different directions; modify amplification; switch between different electromagnet coils; and slide the vibrating membrane. They also can use a compass to measure magnetic fields manually. The system is shown in Figure 1.



**Figure 1.** View of the system without educational AR representations (left) with components labeled; view of the system with educational AR showing magnetic fields, electricity and related representations (right – all holograms and labels are shown as visible in AR).

In (Radu & Schneider, 2019), we have investigated how participants' learning is influenced by augmented reality representations at the individual level (results are summarized in Fig. 3). Participants were randomly assigned to four experimental conditions which comprised a nested design with two factors (primary factor: presence of AR educational representations, sub-factor: type of technology features). All conditions had access to the physical system and were able to interact by changing electricity, generate sound, move the speaker membrane, change the speaker coils, and change amplification; furthermore, all conditions had access to a

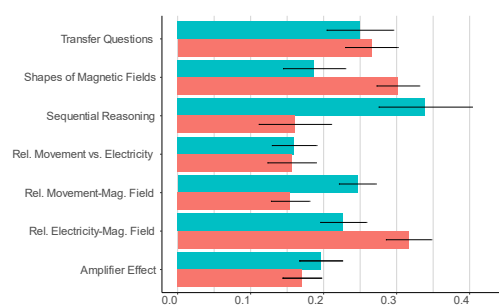


physical poster that explained electromagnetism concepts, had labels showing the function of pieces of the physical system. The groups which did not have educational AR representations (noEdAR) were split into 2 subgroups – dyads which did not wear the HoloLens AR devices, and dyads who did but only saw minimal AR representations composed of holographic labels of major system components, and holograms of sound waves which represented the sound when the system played music. The participants who saw educational AR representations (EdAR) had access to the same information, but could also see interactive visualizations of magnetic fields, electron flows, and electricity graphs. These EdAR groups were split into 2 subgroups in which the presentation of AR visualizations was either presented all at once, or sequenced by a timer. All conditions performed the same study activities.

Our analysis of individual learning gains indicates that participants who used AR technology had significantly higher learning gains on specific concepts such as understanding spatial structures (ex: shapes of magnetic fields through questions, i.e. “Draw the magnetic field around a single wire”), higher ability to transfer knowledge to different situations (ex: transfer questions, i.e. “Is it possible to build a motor that is moved through electric signals? If yes, explain how.”), but were significantly worse at understanding physical effects (ex: questions about relationship between physical movement and magnetic fields, i.e. “If the magnetic field is suddenly inverted, the speaker membrane: a) is pulled closer; b) is pushed away; c) does not move”).



**Figure 2.** A dyad pair interacting with the system while seeing AR through Microsoft HoloLens headsets.



**Figure 3.** Individual differences in relative learning gains (Green: without educational AR; Red: with educational AR; whiskers: standard error).

In the current research we focus on the effects of AR on collaboration and participant leadership and agency, and analyze the data aggregated at the group level. We, compare learning and collaborative behaviors between two conditions: participants who used the system and saw educational AR visualizations; and participants who used the system but did not see educational visualizations. We predict that, due to the increased availability of educational visualization, participants who see educational AR will exhibit less affected by imbalanced leadership and imbalanced knowledge in the group. Our research questions are as follows:

- RQ1: Is overall collaboration and learning impacted by the presence of educational AR?*
- RQ2a: How does participant leadership imbalance impact learning, collaboration, and interaction?*
- RQ2b: How does leadership differ with the presence of AR?*
- RQ3a: How does leadership imbalance impact learning, collaboration, and interaction?*
- RQ3b: How does the effect of driver-follower imbalance differ with the presence of AR?*

## Method

Participants were recruited from the study pool of a laboratory at a university in the northeastern United States. Participation required subjects to not know each other, have no significant prior physics knowledge, be born on/after 1976, speak English fluently, have at least a bachelor’s degree, and wear no bifocal glasses. All participants first completed a pre-test, followed by a short written introduction to relevant physics concepts. Participants then worked on the speaker activity for 30 minutes under different experimental conditions (see Figure 3). During this period, all participants worked on a worksheet and saw a poster of printed physics knowledge on the wall. The study ended with a post-test and debriefing. The variables of interest are as follows:

## Independent variables

We looked at three different independent variables in our analyses: presence / absence of AR, leadership and knowledge imbalance. We provide our operationalization of those constructs below:

**Presence of Educational AR:** Groups were randomly assigned to conditions in which educational AR

was present (EdAR) or not present (NoEdAR), as described above.

**Leadership:** Groups were categorized under two conditions depending on participant leadership: dyads where a leader was present, and dyads where no leader was present. Leadership was considered to be present when both partners in a group do not initiate actions equally (i.e. when the maximum score was not recorded for the qualitatively-observed “reciprocal interaction” dimension, defined in Table 2 below).

**Driver Follower Knowledge:** We rated the participation of each person in a group as either being a “driver” (the participant who typically initiated actions overall) or “follower” (the participant who was more passive overall). We accounted for differences pretest knowledge, resulting in four group configurations: LL, LH, HL, HH where the first letter refers to the driver and the second letter refers to the passenger. For example, HL indicates a high-knowledge driver, and low-knowledge passenger; LH indicates low-knowledge driver and high-knowledge follower.

## Dependent variables

**Learning Metrics:** We measured participant learning through pre- and post-tests. Participants’ learning was compared using relative learning gains, a measure of the relative improvement that occurred between pre-post test scores (Cuendet et. al. 2012). The learning test contained multiple-choice and open-ended questions measuring several aspects of conceptual knowledge. All learning metrics are listed in Figure 3 and described in detail in (Radu & Schneider, 2019).

**Collaboration Metrics:** Collaboration metrics were qualitatively coded for each pair of participants across several dimensions using a validated rating scheme described in Meier, Spada & Rummel (2007). The scale evaluates collaboratives processes through a 5-point scale on the following dimensions: sustained mutual understanding, dialogue management, information pooling, reaching consensus, task division, technical coordination, and reciprocal interaction. Examples are provided in Table 1.

**Interaction Metrics:** The backend of the tangible interface system recorded how much users interacted with different system components, such as changes in electricity direction, movement of speaker membrane, changes of speaker coils, and changes in amplification.

Table 1: The measured dimensions of collaboration, based on Meier, Spada & Rummel (2007)

| Collaboration Metrics          | Example  |
|--------------------------------|--|
| Sustained mutual understanding | Ensure partners are on same page. Speakers make their contributions understandable for their collaboration partner rather than ignoring each other’s insight.  |
| Dialogue management            | Smooth flow of communication with little overlaps. Smooth volley of conversation with little interruptions; partners make sure to have each other’s attention before transitioning to other topics             |
| Information pooling            | Ask questions to seek each other’s perspective. Partners contribute their insight effectively or ask useful questions to seek opinions.  |
| Reaching consensus             | Coming to shared understanding / agreement. Both partners come to a shared conclusion; if there is disagreement, they resolve it through critical consideration  |
| Task division                  | Task division is balanced, and tasks are explicitly distributed between partners through discussion.   |
| Time management                | Deadline met, detailed time planning. Partners monitor the time throughout their cooperation and make sure to finish the current subtask or topic with enough time to complete the remaining subtasks.         |
| Technical coordination         | All tools used, including physical compass. Partners use all technology at their disposal, including features such as physical magnetic compass, using different coils, referring to the physical poster, etc. |
| Reciprocal interaction         | Partners hold equal status in working relationship; both take agency in leading the discussion instead of one partner dominating the working relationship.   |
| Overall Scores                 | Overall Communication (avg. dimensions 1,2), Overall Joint Information Processing (avg. dimensions 3,4), Overall Coordination (avg. dimensions 5,6,7), Overall Collaboration (avg. dimensions 1-8)             |

## Results

In the original study, we recruited 30 participant pairs in the two primary conditions (N=60) and removed sessions where technical issues were encountered or sessions that contained outliers with exemplary pretest knowledge (2.5 standard deviations beyond the pretest mean score), resulting in N=56 sessions (N=28 in EdAR, N=28 in noEdAR).

For qualitative measurement of the degrees of collaboration and leadership, two researchers double-coded 20% of the video recordings and achieved interrater reliability of Cohen Kappa 0.67, indicating “good”

agreement. For statistical tests, normality assumptions were not met for continuous dependent variables, thus we performed Wilcoxon Rank-Sum nonparametric tests for differences between conditions of Presence of Educational AR, and for conditions of Leadership; and we performed Kruskal-Wallis H nonparametric tests followed by post-hoc tests for differences between conditions of Driver-Follower Knowledge.

### RQ1: Are overall collaboration and learning impacted by the presence of educational AR?

We tested for collaboration and learning differences between EdAR and noEdAR conditions. The EdAR groups were significantly better at **time management** ( $W=446, p=0.050$ ), and significantly better at answering **near transfer questions** ( $W=568, p=0.003$ ), but significantly worse at understanding relationship between **magnetic fields and movement** ( $W=248.5, p=0.019$ ). These learning effects mirror our results of analysis at the individual level presented in (Radu & Schneider, 2019), where we found these and additional effects detectable in the larger sample when participants are considered individually (Figure 3). We found no other statistically significant effects when analyzing differences at the dyad level.

### RQ2a: How does participant leadership imbalance impact learning, collaboration, and interaction?

Dyads were categorized into two groups: dyads where a leader was present ( $N=22$  overall;  $N=11$  in EdAR,  $N=11$  in noEdAR), and where no leader was present ( $N=34$  overall;  $N=17$  in EdAR,  $N=17$  in noEdAR).

We analyzed whether there is a significant effect of leadership on dependent variables across all groups. We find that dyads with leadership generally had higher relative learning gains on **all transfer questions** ( $W=253.5, p=0.044$ ); but had weaker collaboration in **sustained mutual understanding** ( $W=511, p=0.010$ ), **reaching consensus** ( $W=584, p<0.001$ ), **task division** ( $W=511, p=0.004$ ), **overall joint information processing** ( $W=562, p=0.001$ ), **overall communication** ( $W=492, p=0.033$ ), and **overall collaboration** ( $W=584.5, p<0.001$ ). This indicates that, although leadership causes some learning gains, it also produces visible negative effects on collaboration.

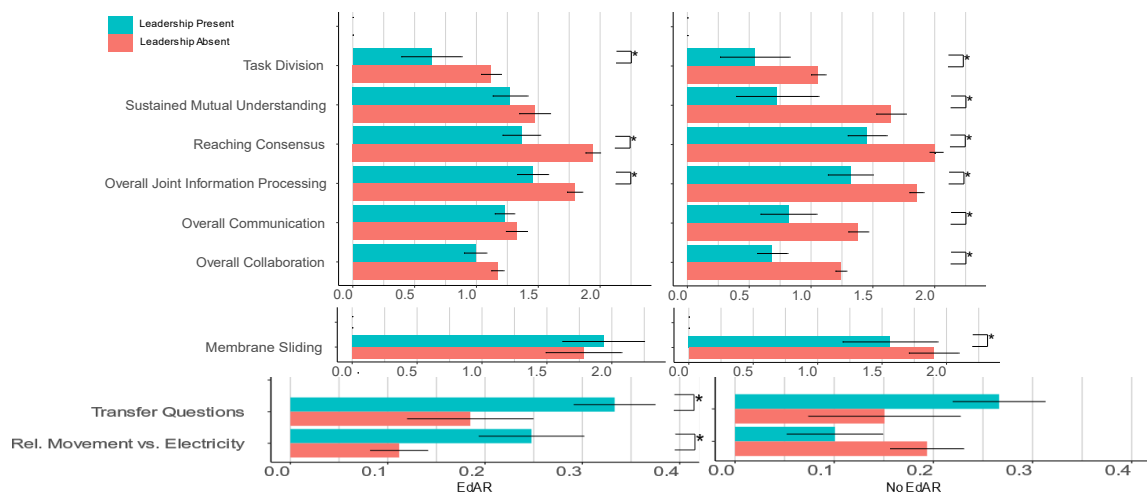


Figure 4. Significant effects of driver-follower configurations in EdAR (left) and noEdAR (right). \*=sig. difference between leadership groups. Whiskers=standard error.

### RQ2b: How does leadership differ with the presence of AR?

Since AR seems to have an influence on group leadership, we investigated the effects of leadership in the presence and absence of educational AR. Within EdAR groups, we found that groups with leadership had higher gains on **transfer questions** ( $W=49.5, p=0.040$ ) and understanding the relationship between **electricity vs. movement** ( $W=47.5, p=0.031$ ); also, groups with leadership had lower scores on **reaching consensus** ( $W=147.5, p=0.001$ ), **task division** ( $W=126, p=0.047$ ), and **overall joint information processing** ( $W=137, p=0.027$ ), compared to groups without leadership. In contrast, in the noEdAR condition, the presence of leadership had no statistical impact on learning gains, and had worse impact on collaboration and system interaction. For groups in the noEdAR condition, similar to EdAR condition, groups with leadership were worse on **reaching consensus** ( $W=144.5, p=0.001$ ), **task division** ( $W=130, p=0.040$ ), **joint information processing** ( $W=144, p=0.009$ ), but had additional negative effects through lower scores on **interaction** with speaker membrane ( $W=82, p=0.039$ ),

**sustained mutual understanding** ( $W=140.5$ ,  $p=0.016$ ), **overall communication** ( $W=138$ ,  $p=0.029$ ), and **overall collaboration** ( $W=162$ ,  $p=0.001$ ). These results indicate that within the EdAR experience, leadership imbalance has a stronger effect of emphasizing learning gains, and has less negative effects on collaboration and system interaction.

To understand the differences between these conditions in relation to leadership effects, we sampled video recordings from EdAR and noEdAR groups. One theme observed is that in situations where EdAR was present, the visual representations helped the more passive participant follow the more dominant participant's explanations: it allowed the passive participants to interrupt and ask for clarification by referencing to the AR representation. Table 2 illustrates an example in EdAR (left) where one participant is teaching the other by using magnetic field polarity representations. In contrast, in the Non-AR example (right) one participant is trying to teach but ends up dominating the discussion while the other participant is unable to keep up with the explanations due to the lack of shared representations.

Table 2: Quotes from participants in moments of teaching

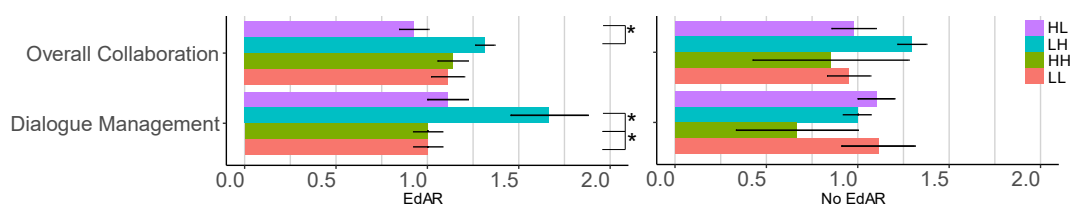
| EdAR Group  | Non EdAR Group  |
|---|---|
| P1: "The direction is, like, the south pole, like N S N S is the backward one." | P2: "So alright here's an interesting thing, when you play music, that's making little currents going forward and backwards, but if no forward current is allowed, it does this <pushes button> this just overpowers everything, and only gonna pull, and when you switch it to forward, only pushes, and when you switch it to music it does little tiny currents like really fast, and then that's making it vibrate, and as it vibrates it's shaking this cup, and as that vibrates it's shaking the air, and that air is shaking our ears, and shaking our brains." |
| P2: "Uh huh."   | P1: "Really cool"   |
| P1: "Can you see that?" <pointing at AR>  | P2: "What do you think?"  |
| P2: "Yeah."   | P1: "Good story" [nervous laughter]   |
| P1: "Starting from the left, N S -"   | P2: "You like my story? Ok, so how does that help us get to the question"   |
| P2: "Yeah."   |   |
| P1: "-N S. And, for forward current, it's -"                                    |   |
| P2: "Oh."   |   |
| P1: "-S N N S."   |   |
| P2: "S N. Ok. So, this is...so backward was [picks up pen]"                     |   |
| P1 [confirms, as P2 draws]: "Backward, N S !"                                   |   |

### RQ3a: How does group leadership imbalance impact learning, collaboration, and interaction?

Each group was categorized into one of the four driver-follower conditions. LL (EdAR  $N=9$ , noEdAR  $N=9$ ), LH (EdAR  $N=6$ , noEdAR  $N=6$ ), HL (EdAR  $N=9$ , noEdAR  $N=10$ ), HH (EdAR  $N=4$ , noEdAR  $N=3$ ). We analyzed the effects of different types of driver-follower configurations across all groups. No statistical differences were found for relative learning gains and interactions. There was a significant effect of driver-follower knowledge on the dimension of **overall coordination** ( $X^2=8.9$ ,  $p=0.031$ ), and **overall collaboration** ( $X^2=11.9$ ,  $p=0.008$ ); in all cases, descriptive LH groups scores higher than other groups. This indicates that overall, groups had better collaboration when novices were guiding the interaction.

### RQ3b: How does the effect of driver-following knowledge differ with the presence of AR?

Since the AR medium may influence how driver-follower configurations behave, we investigated the effects of this variable when educational AR was present or absent. Within the EdAR groups, there were significant effects of driver-follower configurations on the dimensions of **dialogue management** ( $X^2=12.4$ ,  $p=0.006$ ), and **overall collaboration** ( $X^2=10.3$ ,  $p=0.016$ ); in all cases, descriptive statistics show that the LH groups scored higher than other groups. Within the No-EdAR groups, there were no significant effect of driver-follower knowledge distribution. These results suggest that AR educational representations are beneficial to participants in the LH configurations (Fig. 5):



**Figure 5.** Significant effects of driver-follower configurations in EdAR (left) and noEdAR (right). \*=sig. difference between driver-follower configuration conditions. Whiskers=standard error.

By qualitatively analyzing EdAR and noEdAR groups, we saw examples of participants with less knowledge taking initiative in guiding group discussion (Table 3). In groups where AR representations were present, participants with less knowledge could easily ask questions by pointing at hologram representations of the system, and participants with more knowledge had an easier time communicating their knowledge by referring to the visual representations. As an example, Table 3 illustrates how AR representations were useful for guiding participant discussion (left) and lack of such representations in Non-AR groups contributed to difficulties in understanding. In both examples, the participant with less knowledge is asking for clarification from the other participant; in the AR scenario the knowledge is provided by using the AR representations; in the Non-AR scenario, the person with more knowledge is unable to provide a clear explanation. These behaviors appeared to encourage communication within AR groups and may be a reason for increased benefits to dyads in the LH driver-follower configuration.

**Table 3.** Quotes from participants in the AR and non-AR conditions to clarify ambiguous interactions.

| EdAR Group  | Non EdAR Group  |
|---|---|
| P2: "Is it because it's - is it hitting it? Is the -"<br>P1: "I can see the vibration, or I can see how it's moving in this AR [moving hand side to side]"<br>P2: "In the AR?"<br>P1: "AR - the augmented reality part, like it's just - [gesturing around area surrounding cup and coiled wire with hand]"<br>P2: "Yeah"<br>P1: "The angle of the base is like [pointing to cup]"<br>P2: "Oh! Oh! Oh! Yeah, yeah, yeah, yeah, yeah, yeah! I like the - this thing [moving finger back and forth in front of area between cup and coiled wire]. Okay, so the alternating currents is pulling the magnet closer or further - further"<br>P1: "Mm-hm"<br>P2: "Okay" | P1: "So, when it, when it's a backward current [illustrating shape of magnetic field using hands], it looks like it's pulling this way"<br>P2 [pointing toward cup]: "On this, this thing, right?"<br>P1: "Yeah, so I wonder - I wonder if it's almost like, when it's backward, it's like that [illustrating on paper with fingers]."<br>P2 [confused, handing pen over]: "Are you good at drawing?"<br>P1 [starting to draw on worksheet]: "Well, I would, um..." |

## Discussion

Our original hypotheses predicted that the presence of educational AR representations would have an effect on overall collaboration and learning. For collaboration, we found differences in time management, with lower values for groups which did not have educational AR representations and tended to run out of time. This indicates that the AR representations were useful for completing the task, and groups which lacked these had a harder time understanding the system and communicating since they had to generate their own representations (for example by drawing). For learning gains, analyses at dyad level did not show the same effect as the individual analysis presented in (Radu & Schneider, 2019). This is likely due to the effect of averaging the learning between a participant pair, which reduced the sample size and decreased our statistical power.

Augmented reality was also helpful for groups where participants had knowledge imbalance, specifically when the participants with high knowledge took a back seat and allowed the low-knowledge participants to drive the interaction. When this occurred, metrics of group collaboration was higher in AR groups. This indicates that AR can potentially improve collaboration in groups where participants have unequal knowledge. This might be due to the availability of information to both participants, which can equalize the effect of knowledge imbalance by allowing less-knowledgeable participants to more easily communicate points of confusion by referring to the existing AR representation. Referring to the present representations might also allow the more knowledgeable participants to teach his/her peer, which can help passive participants be more engaged in the collaborative activity.

These results point at potential future work investigating the detrimental effects of augmented reality on learning concepts focused on physicality (ex: relationship between movement and magnetic fields), that possibly may be caused by learners being hyper-focused by highly visual nature of the AR medium and paying less attention to physical sensations. On the other hand, these results indicate that AR can be beneficial for equalizing the effects of imbalanced collaboration, and future work could investigate how this can benefit situations where participants tend to be imbalanced in agency (ex: in student team projects where some participants are more high-achieving than others) or in knowledge (ex: in teams of varied expertise levels, where participants may need to take on the role of teachers).

## Conclusions

In this study we analyzed collaboration, learning and interactions of dyad pairs as they experienced an AR system for learning about electromagnetism. We found that, in this context, augmented reality was generally beneficial for both learning and collaboration. Overall, AR representations improved time management, learning of structural concepts but reduced learning of physical concepts. The effects of leadership were mediated by the presence of AR: the presence of leaders in AR was linked to higher learning gains and better collaboration, whereas these effects were not present without AR. Finally, for groups with imbalanced knowledge, AR seemed to benefit participants in configurations where less-knowledgeable participants took the lead in discussions.

## References

- Belcher, J. W., & Bessette, R. M. (2001). MIT educators share success-Using 3D animation in teaching introductory electromagnetism. *COMPUTER GRAPHICS-US*, 35(1), 18–21.
- Billinghurst, M., Clark, A., & Lee, G. (2015). A survey of augmented reality. *Foundations and Trends® in Human-Computer Interaction*, 8(2–3), 73–272.
- Chang, Y.-L., Hou, H.-T., Pan, C.-Y., Sung, Y.-T., & Chang, K.-E. (2015). Apply an augmented reality in a mobile guidance to increase sense of place for heritage places. *Journal of Educational Technology & Society*, 18(2).
- Chen, Y.-C. (2006). A study of comparing the use of augmented reality and physical models in chemistry education. In *Proceedings of the 2006 ACM international conference on Virtual reality continuum and its applications* (pp. 369–372). ACM.
- Church, T., Hazlewood, W. R., & Rogers, Y. (2006). *Around the table: Studies in co-located collaboration*. Austrian Computer Society.
- Cuendet, S., Bumbacher, E., & Dillenbourg, P. (2012). Tangible vs. virtual representations: when tangibles benefit the training of spatial skills. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design* (pp. 99–108). ACM.
- Dunleavy, M., & Dede, C. (2014). Augmented reality teaching and learning. In *Handbook of research on educational communications and technology* (pp. 735–745). Springer.
- Ibáñez, M. B., Di Serio, Á., Villarán, D., & Kloos, C. D. (2014). Experimenting with electromagnetism using augmented reality: Impact on flow student experience and educational effectiveness. *Computers & Education*, 71, 1–13.
- Maloney, D. P., O’Kuma, T. L., Hieggelke, C. J., & Van Heuvelen, A. (2001). Surveying students’ conceptual knowledge of electricity and magnetism. *American Journal of Physics*, 69(S1), S12–S23.
- Meier, A., Spada, H., & Rummel, N. (2007). A rating scheme for assessing the quality of computer-supported collaboration processes. *International Journal of Computer-Supported Collaborative Learning*, 2, 63–86.
- Morrison, A., Oulasvirta, A., Peltonen, P., Lemmela, S., Jacucci, G., Reitmayr, G., ... Juustila, A. (2009). Like bees around the hive: a comparative study of a mobile augmented reality map. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1889–1898). ACM.
- Phon, D. N. E., Ali, M. B., & Halim, N. D. A. (2014). Collaborative augmented reality in education: A review. In *Teaching and Learning in Computing and Engineering (LaTiCE), 2014 International Conference on* (pp. 78–83). IEEE.
- Radu, I., Doherty, E., DiQuollo, K., McCarthy, B., & Tiu, M. (2015). Cyberchase shape quest: pushing geometry education boundaries with augmented reality. In *Proceedings of the 14th International Conference on Interaction Design and Children* (pp. 430–433). ACM.
- Radu, I. & Schneider, B. (2019). What Can We Learn from Augmented Reality (AR)? Benefits and Drawbacks of AR for Inquiry-based Learning of Physics. In *2019 CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland, UK*. ACM.
- Radu, I. (2014). Augmented reality in education: a meta-review and cross-media analysis. *Personal and Ubiquitous Computing*, 18(6), 1533–1543.
- Salomon, G., & Globerson, T. (1989). When teams do not function the way they ought to. *International Journal of Educational Research*, 13(1), 89–99.
- Yuill, N., & Rogers, Y. (2012). Mechanisms for collaboration: A design and evaluation framework for multi-user interfaces. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 19(1), 1.

## Acknowledgements

This material is based on work supported by the National Science Foundation, Grant No. 1748093. We thank all our participants, reviewers, the Harvard Decision Science Lab, and assistants who helped with data collection.

# Finding Common Ground: A Method for Measuring Recent Temporal Context in Analyses of Complex, Collaborative Thinking

Andrew R. Ruis, University of Wisconsin–Madison, arruis@wisc.edu  
Amanda L. Siebert-Evenstone, University of Wisconsin–Madison, alevenstone@wisc.edu  
Rebecca Pozen, University of Wisconsin–Madison, rmpozen@gmail.com  
Brendan R. Eagan, University of Wisconsin–Madison, beagan@wisc.edu  
David Williamson Shaffer, University of Wisconsin–Madison & Aalborg University, david.shaffer@wisc.edu

**Abstract:** Complex, collaborative thinking is often conceptualized as a process of developing cognitive connections among the contributions of different participants. A central problem in modeling collaboration in this way is thus determining, for any contribution to a discussion, the appropriate context for modeling the connections being made—that is, for determining the appropriate recent temporal context. Recent temporal context is typically defined using a moving window of fixed length. However, that length is dependent on the setting, and there are no existing methods for reliably determining an appropriate window length. This paper presents an empirical method for measuring recent temporal context, and thus for defining an appropriate window length to be used in analyses of complex, collaborative thinking. Importantly, the method we describe minimizes the need for human annotation while providing both qualitative and quantitative warrants for choosing a particular window length.

## Introduction

In the learning sciences, complex thinking is often conceptualized as a process of developing cognitive connections among concepts (DiSessa, 1988; Linn, Eylon, & Davis, 2004; Shaffer, 2012). In computer-supported collaborative learning (CSCL) contexts, individuals make such connections not only within their own contributions but also to the contributions of their collaborators (Garrison, Anderson, & Archer, 2001; Shaffer, 2017). A central problem in modeling complex, collaborative thinking in terms of cognitive connections is thus determining, for any contribution to a discussion, the appropriate temporal context for modeling the connections being made. Prior work in the learning sciences has approached this problem using *moving windows* (Dyke, Rohit Kumar, Hua, & Rosé, 2012; Siebert-Evenstone et al., 2016), where each turn of talk is associated with some prior segment of the discussion that forms its *recent temporal context* (Suthers & Desiato, 2012).

In such models, analysis of a given turn of talk accounts for both its own content and the content of its associated window. Because there are many variables that may affect the extent of the recent temporal context, including domain, topical complexity, age of the participants, and communication medium, it is important to identify an appropriate window length for each setting. However, there are no reliable methods for measuring the extent of recent temporal context, and studies have not been conducted that show the effects of window length on the features or interpretation of window-based learning analytic models. In this study, we present an empirical method for measuring the extent of recent temporal context. We then evaluate the method by analyzing conceptual connectivity in the same dataset using different window lengths to explore the effects of window length on the resulting models. The results suggest that an appropriate moving window length can be empirically determined with minimal effort, and that while window length can significantly affect model features and interpretation, the empirical method we describe produces relatively robust models of complex thinking.

## Background

Collaborative learning takes place both in face-to-face interactions and remotely, mediated by various communications technologies. In addition, computer simulations, intelligent tutoring systems, educational games, and other CSCL learning technologies create settings in which students work with one another, with educators or mentors, and with pedagogical agents to frame, investigate, and solve complex problems. Such settings foster the development of *communities of inquiry* (Garrison & Akyol, 2013; Seixas, 1993): groups of participants who interact with one another in a problem-solving context to facilitate critical discourse and establish shared meaning and understanding. A key goal of such collaborative interactions is to help participants make explicit connections among the contributions of different team members, which enables communities of inquiry to organize knowledge, identify and address misconceptions, facilitate reflection, and ultimately construct mutual understanding (Garrison et al., 2001).

Modeling complex, collaborative thinking, then, requires assessing not only the connections that individuals make among their own contributions but also the connections they make to the contributions of the other participants. This in turn requires identifying the appropriate relational context for any given turn of talk in a discussion (Arvaja, Salovaara, Häkkinen, & Järvelä, 2007). Researchers often establish relational context by looking at each turn of talk and identifying its *referents*: the prior turn or turns of talk that provide the information needed to understand the meaning of a given utterance (Shaffer, 2017).

### Illustrating the problem

Consider the following excerpt, in which a student project team working on a biomedical engineering design project is discussing with a mentor (Maria) how they determined which design prototypes to test (emphasis added).

- 1 Maria What role did the requests of the **internal consultants** play in deciding which prototypes to test?
- 2 Jill that's it though. then we have to pick a final one and make a poster presentation
- 3 Jill **Internal consultants** made it clear **reactivity** is important, but **cost** triumphs
- 4 Lesenia the **requests** gave a focus on what **factors** to pay attention to
- 5 Maria How did you design your device to address **patient** needs?
- 6 William We tried to make them **cheap** and **reliable** so they could last a long time and save the **patients** money
- 7 Brad Since we are trying to **make a bunch of people happy**, we wanted a wide **variety of results** to see which one **makes the most people happy**.

Maria asks the students a question in Line 1, which Jill and Lesenia answer in Lines 3 and 4. Maria then asks a second question in Line 5, to which William responds in Line 6. In Line 7, Brad's response references "people" twice, and we can identify who they are only by looking to previous utterances—that is, by identifying the referent(s) for Line 7. Yet how far back do we need to look? If the context for his utterance consists only of Lines 5 and 6, we may determine that Brad is referring to patients, the users of the device the team is designing. However, if all the lines are included as context, it is clear that by "people" Brad means both the patients *and* the internal consultants. Interpretation of Brad's utterance—and by extension, interpretation of the connections that Brad is contributing to the team's mutual understanding—depends on how we define his contribution's recent temporal context. That is, there is some window (in this case, seven turns of talk) that contains all of the context (i.e., all of the referents) needed to understand a given contribution to a discussion.

Ideally, the size of a given window would be determined by the content of the discussion, as the context needed for interpreting any given turn of talk will vary. In the excerpt above, understanding Brad's contribution requires a window of seven turns of talk, while understanding William's contribution in Line 6 requires only two: he is answering Maria's question in Line 5, while Brad is addressing both of Maria's questions (Lines 1 and 5) and building on the contributions of Jill, Lesenia, and William. Note that if a larger window were used to analyze William's contribution, he would be credited for making connections to the internal consultants and their concerns, which he does not seem to be doing. In other words, a larger window may produce more false positive connections. However, while humans can identify relational context with relative ease, manual annotation of discourse is not a scalable approach. As large volumes of rich process data are increasingly produced by a variety of CSCL environments, rapid and reliable procedures for measuring recent temporal context are necessary for analyses of complex, collaborative thinking at scale.

### Existing approaches

Research on natural language processing suggests that identifying changes in topic or turn-by-turn references to prior utterances cannot be reliably automated (Rosé et al., 2008). Even if automated processes were feasible, the excerpt above illustrates a key problem with segmentation by changes in topic. In Line 5, Maria *is* changing the topic—from the needs of the internal consultants to the needs of the patients—but Brad (Line 7) uses that as an opportunity to make a point about how, as engineers, the project team needs to consider the concerns of multiple different stakeholders when designing products. That is, he is making a connection between the two topics (Lines 1–4 and Lines 5–6). Moreover, even if a given utterance does not make explicit reference to any of the immediately preceding turns of talk, those utterances nonetheless form the *common ground* for that part of the conversation and are likely to influence what participants say and do (Clark, 1996).



Thus for both practical and theoretical reasons, researchers typically use a moving window of fixed length (e.g., a particular duration of time or number of turns of talk) to analyze collaborative connection-making in recent temporal context. Instead of creating summary values for all utterances in a conversation, a moving window analysis computes a value for each utterance, based on the content of that utterance and the content of preceding utterances that are contained within the window. This type of analysis has been used to explore intragroup interactions (Dyke et al., 2012) and to study connection-making in collaborative discourse (Csanadi, Eagan, Shaffer, Kollar, & Fischer, 2017; Quardokus Fisher, Hirshfield, Siebert-Evenstone, Arastoopour, & Koretsky, 2016; Siebert-Evenstone et al., 2016; Sullivan et al., 2018; Suthers & Desiato, 2012).

One technique that uses moving windows to model connections in recent temporal context is *epistemic network analysis* (ENA) (Shaffer, 2018; Shaffer, Collier, & Ruis, 2016; Shaffer & Ruis, 2017). ENA takes interaction data coded for elements of complex thinking and constructs weighted network models that represent the structure of connections made among those elements by each individual in the dataset. For each line in a conversation, ENA computes a network of connections within the recent temporal context as defined by the window length; that is, ENA identifies the co-occurrences of codes between a given utterance and the previous utterances in the conversation that fall within the window. Proceeding line-by-line through the data, ENA accumulates these networks for each individual to model the unique connections contributed by that person to the conversation. Critically, the length of the moving window—in ENA and in other window-based techniques for modeling conceptual connectivity—defines which *co-occurrences of codes* are hypothesized to represent *meaningful cognitive connections*.

## The challenge

A key challenge for window-based models of complex, collaborative thinking is thus to determine a fixed window length that is sufficiently long to capture the recent temporal context for most utterances but not so long as to overrepresent connections that are not meaningful. Because the length of the window determines which connections are included in the model, interpretation of the model may be significantly affected by the choice of window length (Gleicher, 2016). Researchers typically make such determinations on a case-by-case basis, usually without detailing the method used, and an extensive literature search revealed no scalable, objective techniques for measuring recent temporal context. In what follows, we present a novel method for determining window length that minimizes the need for human annotation and provides both qualitative and quantitative warrants for using a certain window length to analyze collaborative connection-making in the specified setting.

## Methods

### Setting and data

In this study, we analyzed the collaborative interactions of students in the engineering virtual internship *Nephrotex* (Arastoopour, Shaffer, Swiecki, Ruis, & Chesler, 2016; Chesler, Arastoopour, D'Angelo, Bagley, & Shaffer, 2013; Chesler et al., 2015). In *Nephrotex*, students work in project teams to design an ultrafiltration membrane for a hemodialysis system. The virtual internship is divided into a series of activities that simulate steps in the design process, including reviewing research reports; designing prototypes; discussing design choices with teammates and an engineering advisor; and addressing the needs of internal consultants and external clients. Students interact with their teams and with their engineering advisor through an online instant message program (chat), and the system automatically records all chat conversations for subsequent analysis. *Nephrotex* takes approximately 15 hours to complete.

Chat conversations ( $N = 54,896$  chats) were collected from 20 runs of *Nephrotex* at five institutions in the United States. Participants ( $N = 652$ ) were first- and second-year college students using *Nephrotex* as part of an engineering course. Students were randomly assigned to project teams of 4–5 members.

### Window length annotation

To measure recent temporal context in this setting, we randomly selected 200 utterances from the 54,896 chats in the *Nephrotex* dataset. For each chat, two independent raters identified all previous referents in the conversation. These annotations indicate, for each utterance, the window containing that utterance's recent temporal context, where window length is the number of chats from the referring utterance to the earliest referent, inclusive. In the excerpt above, for example, the window length for Brad's utterance is seven lines: the referring chat (Line 7); the furthest referent (Maria's chat in Line 1); and the intervening five chats.

To calculate agreement between the two independent raters, we computed Cohen's  $\kappa$  (kappa) for each window length. Kappa was calculated for each window size,  $x$ , by assigning a "1" to any utterance that a given rater determined to have window length  $x$  and a "0" to all other utterances. Kappa thus indicated the extent to

which the two raters agreed in their assessments of each utterance’s recent temporal context. To determine whether kappa scores for the sample (200 chats) could be generalized to the population from which they were drawn (> 50,000 chats), we computed Shaffer’s  $\rho$  (rho) to estimate the expected Type I error rate of kappa given the sample size (Eagan, Rogers, Pozen, Marquart, & Shaffer, 2016; Shaffer, 2017).

### Effects of window length on models of connectivity

We tested our technique for empirically determining window length by analyzing a portion of the *Nephrotex* dataset using ENA (for a detailed description of ENA methodology, see Shaffer, Collier, & Ruis, 2016). Specifically, we used ENA to model data from two implementations of *Nephrotex* (48 students; 5,757 chats) at window length  $x$  for each  $x \in \{1, 2, \dots, 13\}$ . At each window length, we compared the networks of (a) students using an engineering virtual internship for the first time (novices;  $n = 24$ ), and (b) students using *Nephrotex* after using a different engineering virtual internship (relative experts;  $n = 24$ ).

The data were coded using an automated coding algorithm for five elements of engineering design: (1) *performance parameters*: the functional attributes of a design; (2) *design decisions*: the process of making design choices, including prioritization and tradeoffs; (3) *client requests*: considering the concerns and needs of stakeholders; (4) *data*: considering technical or numeric information; and (5) *collaboration*: facilitating inclusivity and teamwork in the design process. Inter-rater reliability was computed for each code separately, and all codes had a kappa value greater than 0.75; each kappa value was statistically significant at  $\alpha < 0.05$  and  $\kappa > 0.65$ :  $N = 200, \kappa > 0.75, \rho(0.65) < 0.05$ .

## Results

### Measurement of recent temporal context

Of the 200 chats examined, 49 (24.5%) made no reference to prior chats, and 51 (25.5%) referenced only the previous chat. Thus, a window length of two chats would capture the relevant connections for 50% of the utterances in the sample. However, as Table 1 shows, it is not until a window length of seven that the relevant connections were captured for more than 95% of the utterances. Subsequent increases in window length resulted in only very small improvements, and no utterance required a window length of more than 18 chats.

**Table 1: Number and proportion of utterances with complete recent temporal context at each window length**

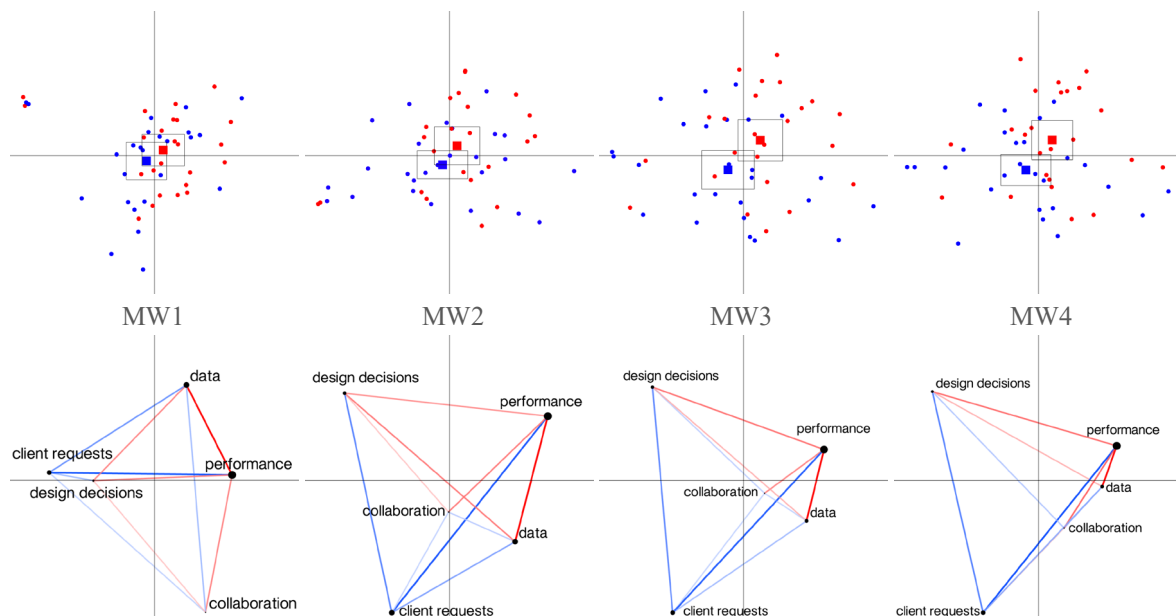
| Window Length | Number (and Percentage) of Utterances with Complete RTC | Increase in Percentage of Utterances with Complete RTC over Previous Window Length | Cohen’s $\kappa$ | Shaffer’s $\rho$ (for $\kappa > 0.65$ ) |
|---------------|---|--|------------------|---|
| 1             | 49 (24.5%)  |  | 0.96             | < <b>0.01**</b>                         |
| 2             | 100 (50.0%)   | +25.5%   | 0.95             | < <b>0.01**</b>                         |
| 3             | 131 (65.5%)   | +15.5%   | 0.96             | < <b>0.01**</b>                         |
| 4             | 158 (79.0%)   | +13.5%   | 0.97             | < <b>0.01**</b>                         |
| 5             | 170 (85.0%)   | +6.0%  | 0.88             | < <b>0.01**</b>                         |
| 6             | 182 (91.0%)   | +6.0%  | 0.84             | < <b>0.01**</b>                         |
| <b>7</b>      | <b>192 (96.0%)</b>                                      | <b>+5.0%</b>   | <b>0.94</b>      | <b>&lt; 0.01**</b>                      |
| 8             | 195 (97.5%)   | +1.5%  | 0.91             | <b>0.03*</b>                            |
| 9             | 195 (97.5%)   | +0.0%  | 0.91             | <b>0.02*</b>                            |
| 10            | 197 (98.5%)   | +1.0%  | 0.85             | 0.11                                    |
| 11            | 197 (98.5%)   | +0.0%  | 0.66             | 0.39                                    |
| 12            | 198 (99.0%)   | +0.5%  | 0.80             | 0.22                                    |
| 13            | 198 (99.0%)   | +0.0%  | 0.80             | 0.22                                    |

To calculate the level of agreement between the two raters and to determine whether the findings for the sample (200 chats) are generalizable to the population from which the chats were sampled (> 50,000 chats), we computed kappa and rho for each window length (see Table 1). For all window lengths up to nine, agreement between the two raters was statistically significant for  $\kappa > 0.65$  ( $N = 200, \kappa \geq 0.84, \rho(0.65) < 0.05$ ), which indicates that the level of agreement between the two raters would have been  $\kappa > 0.65$  for those window lengths had they evaluated the entire dataset. While the kappa scores obtained for window lengths greater than nine were not

statistically significant at a Type I error rate of  $\alpha < 0.05$  (i.e.,  $p(0.65) > 0.05$ ), this is due to the extremely low number of utterances that refer to chats more than 8 turns earlier. Because the goal of this method is to identify the shortest window that captures the full recent temporal context for most utterances, a window length of seven appears to be optimal for this dataset.

### Effects of window length on model features and interpretation

Figure 1 shows the ENA models for moving window lengths one through four (MW1–4). The plotted points (top row), each of which represents the network of one student, show increasing discrimination between novices (red) and relative experts (blue) with increasing window length. The difference graphs (bottom row) compute the difference in connection strengths between the mean networks of the novices and relative experts, showing which connections were stronger in which group. Note that the network graph for MW1, which includes only connections made within individual utterances, is very different from the network graph for MW2, which includes both connections within a given utterance and connections to the immediately preceding utterance. With increasing window length, the network graphs become more stable, as does the interpretation of the model. As the difference graphs for MW3 and MW4 indicate, the novice students made stronger connections among *data*, *performance*, and *design decisions*, while the relative expert students made stronger connections between *client requests* and both *performance* and *design decisions*. In other words, the models for those window lengths suggest that novices were more likely to focus on the technical aspects of the design problem, while relative experts were more likely to focus on the needs of stakeholders in the design process.



**Figure 1.** ENA models of the same *Nephrotex* data at four moving window (MW) lengths. The plotted points (top row) show the network locations of the novices (red) and relative experts (blue), with the corresponding means (colored squares) and 95% confidence intervals (boxes). The difference graphs (bottom row) show the difference in connection strength between the means of the two groups.

The summary statistics for the ENA models at all window lengths tested (MW1–13) are shown in Table 2. With the exception of MW5, all ENA models produced using a moving window larger than two indicate a significant difference ( $p < 0.05$ ) between novices and relative experts with a moderate-to-large effect size ( $d > 0.60$ ). However, at a window length of seven, which was hypothesized to be optimal based on our qualitative analysis, the difference between the two groups is significant at  $p < 0.01$  for the first time, and the effect size increases almost 10% from the window length of six. The effect size increases further at MW8, but all window lengths from MW9 to MW13 produce summary statistics comparable to MW7.

Figure 2 shows the ENA model for MW7. Note that the both the discrimination between the groups (left) and the structure of connections (right) are similar to the MW4 model shown in Figure 1. While the ability of ENA to discriminate between novices and relative experts is relatively robust to window length, interpretation of the model depends in part on the position of the network nodes in ENA space. The nodes are positioned based on the patterns of connectivity, which in turn are affected by changes in window size. Thus, differences in node

positions between models with different window sizes indicate potentially different interpretations of complex, collaborative thinking (1).

Table 2: ENA model discrimination between two study populations at each window length

| Window Length | $t$  | $p$                | Cohen's $d$ |
|---------------|------|--------------------|-------------|
| 1             | 0.91 | 0.37               | 0.26        |
| 2             | 1.58 | 0.12               | 0.45        |
| 3             | 2.15 | <b>0.04*</b>       | 0.62        |
| 4             | 2.46 | <b>0.02*</b>       | 0.72        |
| 5             | 1.55 | 0.13               | 0.45        |
| 6             | 2.44 | <b>0.02*</b>       | 0.71        |
| 7             | 2.75 | <b>&lt; 0.01**</b> | 0.79        |
| 8             | 3.01 | <b>&lt; 0.01**</b> | 0.87        |
| 9             | 2.83 | <b>&lt; 0.01**</b> | 0.82        |
| 10            | 2.75 | <b>&lt; 0.01**</b> | 0.80        |
| 11            | 2.71 | <b>&lt; 0.01**</b> | 0.79        |
| 12            | 2.68 | <b>0.01*</b>       | 0.78        |
| 13            | 2.70 | <b>0.01*</b>       | 0.79        |

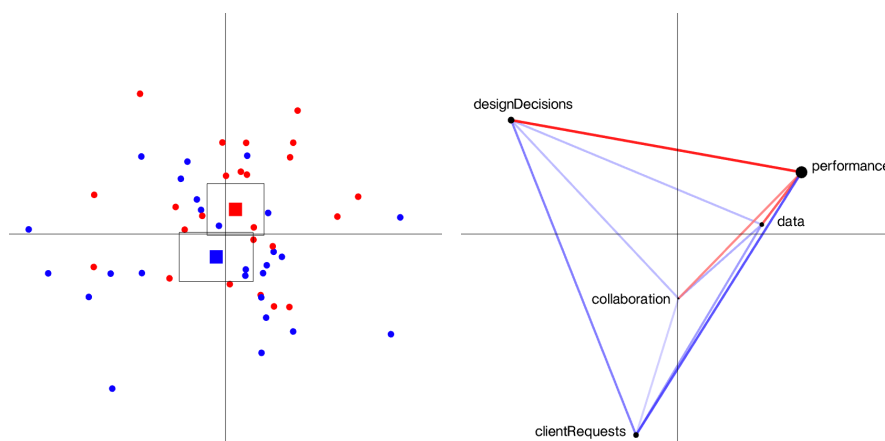


Figure 2. ENA model of the *Nephrotex* data at MW7.

To assess at what window length the network node positions stabilize, we plotted the locations of the nodes at each window length for both the first ( $x$ ) and second ( $y$ ) dimensions of the ENA model (see Figure 3).

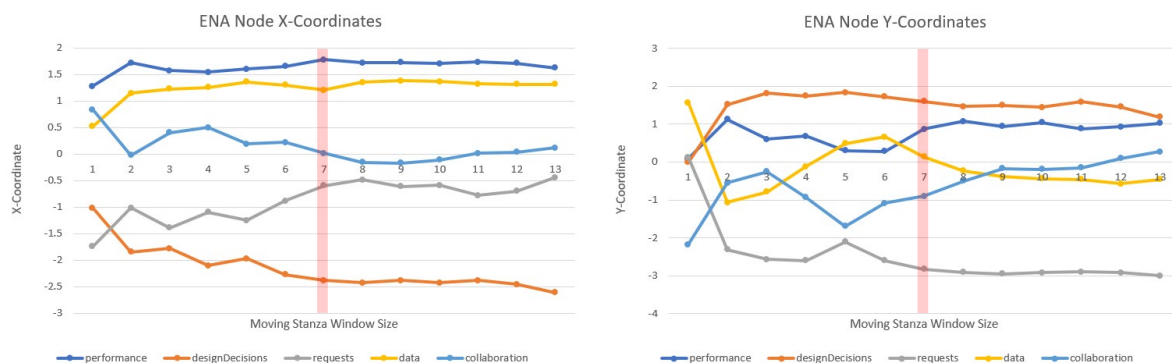


Figure 3. Node positions ( $x$  and  $y$  coordinates) in the ENA models at each window length.

The  $x$ -coordinates of the nodes maintain the same relative order starting at MW2, and the relative spacing stabilizes starting at MW7. After MW7, there are no significant changes in either the relative order or the relative

spacing of the nodes. However, the relative order of the  $y$ -coordinates of the nodes does not begin to stabilize until MW7. *Data* (yellow) and *collaboration* (light blue) change position relative to one another between MW8 and MW9, but the difference is small and both nodes remain near the origin; thus, the change does not affect interpretation of model. A moving window length of seven is thus where the node positions in ENA space stabilize on both dimensions. Window lengths greater than seven do not add significant information to the model, and window lengths less than seven show different patterns of connectivity, even though models using smaller window lengths do discriminate between the novice and relative expert students.

## Discussion

Our goal was to develop a method that (a) provides both qualitative and quantitative warrants for determining the optimal window length for use in moving window analyses of conceptual connectivity in CSCL contexts, while (b) minimizing the number of items requiring human evaluation. To assess this approach, two independent raters analyzed a random sample of 200 student chats (< 0.01% of the 54,896 chats in the dataset). This method identified MW7 as the optimal window length for analyzing these data. We then constructed ENA models of the data that differed only in the choice of window length. This analysis confirmed that a model with a window size of 7 both (a) provides statistical discrimination between groups known to exhibit different patterns of conceptual connectivity based on prior research (Chesler et al., 2015) and (b) provides a stable interpretation of the ENA model. This analysis suggests that statistical discrimination between two groups may be fairly robust to window length once some minimum length is reached (in this case, after MW3 with the exception of MW5), but that model features and interpretation may be more sensitive to window length (the features and interpretation of the ENA models do not fully stabilize until MW7).

As a result, we argue that annotating a subset of data for *furthest referents* makes it possible to analyze recent temporal context and thus determine an appropriate window length to be used in analyses of complex, collaborative thinking. Importantly, this method minimizes the need for human annotation while providing both qualitative and quantitative warrants for choosing a particular window length. While we describe this method by presenting results from one CSCL context (*Nephrotex*) and one learning analytic technique (ENA), we believe that approaches based on annotating a sample data for furthest referents will be compatible with different CSCL settings, different theories of collaborative discourse, and different methods for modeling conceptual connectivity using moving windows. Of course, future research should test our method by repeating this study using other data and other learning analytic models.

While there are many avenues for additional research, this study suggests that hand annotation of a relatively small number of utterances can be used to measure recent temporal context. Critically, this method provides a warrant for making generalizations to the population from which the hand-annotated sample was drawn, making it suitable for analyses of complex, collaborative thinking at scale.

## Endnotes

- (1) The goodness of fit measures are high (both Spearman's and Pearson's  $r > 0.95$ ) for all 13 models except MW1 ( $r \geq 0.85$ ), which means that the relative positions of the network nodes provide good interpretations of the differences between networks in the model.

## References

- Arastoopour, G., Shaffer, D. W., Swiecki, Z., Ruis, A. R., & Chesler, N. C. (2016). Teaching and assessing engineering design thinking with virtual internships and epistemic network analysis. *International Journal of Engineering Education*, 32(3B), 1492–1501.
- Arvaja, M., Salovaara, H., Häkkinen, P., & Järvelä, S. (2007). Combining individual and group-level perspectives for studying collaborative knowledge construction in context. *Learning and Instruction*, 17(4), 448–459.
- Chesler, N. C., Arastoopour, G., D'Angelo, C. M., Bagley, E. A., & Shaffer, D. W. (2013). Design of a professional practice simulator for educating and motivating first-year engineering students. *Advances in Engineering Education*, 3(3), 1–29.
- Chesler, N. C., Ruis, A. R., Collier, W., Swiecki, Z., Arastoopour, G., & Shaffer, D. W. (2015). A novel paradigm for engineering education: Virtual internships with individualized mentoring and assessment of engineering thinking. *Journal of Biomechanical Engineering*, 137(2), 024701:1-8.
- Clark, H. H. (1996). *Using language*. Cambridge University Press.
- Csanadi, A., Eagan, B., Shaffer, D. W., Kollar, I., & Fischer, F. (2017). Collaborative and individual scientific reasoning of pre-service teachers: New insights through epistemic network analysis (ENA). In B. K. Smith, M. Borge, E. Mercier, & K. Y. Lim (Eds.), *Making a difference: Prioritizing equity and access*

- in *CSCL: 12th International Conference on Computer-Supported Collaborative Learning* (Vol. I, pp. 215–222).
- DiSessa, A. A. (1988). Knowledge in pieces. In G. Forman & P. Pufall (Eds.), *Constructivism in the computer age* (pp. 47–70). Hillsdale, NJ: Erlbaum.
- Dyke, G., Rohit Kumar, R., Hua, A., & Rose, C. P. (2012). Challenging assumptions: Using sliding window visualizations to reveal time-based irregularities in CSCL processes. In *Proceedings of the 10th International Conference of the Learning Sciences* (pp. 363–370).
- Eagan, B. R., Rogers, B., Pozen, R., Marquart, C., & Shaffer, D. W. (2016). rhoR: Rho for inter rater reliability (Version 1.1.0). Retrieved from <https://cran.r-project.org/web/packages/rhoR/index.html>
- Garrison, D. R., & Akyol, Z. (2013). The community of inquiry theoretical framework. In M. G. Moore (Ed.), *Handbook of distance education* (Vol. 3, pp. 104–120). Routledge.
- Garrison, D. R., Anderson, T., & Archer, W. (2001). Critical thinking, cognitive presence, and computer conferencing in distance education. *American Journal of Distance Education*, 15(1), 7–23.
- Gleicher, M. (2016). A framework for considering comprehensibility in modeling. *Big Data*, 4(2), 75–88.
- Linn, M. C., Eylon, B.-S., & Davis, E. A. (2004). The knowledge integration perspective on learning. In M. C. Linn, E. A. Davis, & P. Bell (Eds.), *Internet environments for science education* (pp. 29–46). Mahwah, NJ: Lawrence Erlbaum Associates.
- Quardokus Fisher, K., Hirshfield, L., Siebert-Evenstone, A. L., Arastoopour, G., & Koretsky, M. (2016). Network analysis of interactions between students and an instructor during design meetings. In *Proceedings of the American Society for Engineering Education* (p. 17035). ASEE.
- Rosé, C., Wang, Y.-C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., & Fischer, F. (2008). Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *International Journal of Computer-Supported Collaborative Learning*, 3(3), 237–271.
- Seixas, P. (1993). The community of inquiry as a basis for knowledge and learning: The case of history. *American Educational Research Journal*, 30(2), 305–324.
- Shaffer, D. W. (2012). Models of situated action: Computer games and the problem of transfer. In C. Steinkuehler, K. D. Squire, & S. A. Barab (Eds.), *Games, learning, and society: Learning and meaning in the digital age* (pp. 403–431). Cambridge, UK: Cambridge University Press.
- Shaffer, D. W. (2017). *Quantitative ethnography*. Madison, WI: Cathcart Press.
- Shaffer, D. W. (2018). Epistemic network analysis: Understanding learning by using big data for thick description. In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, & P. Reimann (Eds.), *International handbook of the learning sciences* (pp. 520–531). New York, NY: Routledge.
- Shaffer, D. W., Collier, W., & Ruis, A. R. (2016). A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics*, 3(3), 9–45.
- Shaffer, D. W., & Ruis, A. R. (2017). Epistemic network analysis: A worked example of theory-based learning analytics. In C. Lang, G. Siemens, A. F. Wise, & D. Gasevic (Eds.), *Handbook of learning analytics* (pp. 175–187). Society for Learning Analytics Research.
- Siebert-Evenstone, A. L., Arastoopour, G., Collier, W., Swiecki, Z., Ruis, A. R., & Shaffer, D. W. (2016). In search of conversational grain size: Modeling semantic structure using moving stanza windows. In C.-K. Looi, J. Polman, U. Cress, & P. Reimann (Eds.), *Transforming learning, empowering learners: The International Conference of the Learning Sciences (ICLS) 2016* (Vol. I, pp. 631–638).
- Sullivan, S. A., Warner-Hillard, C., Eagan, B. R., Thompson, R., Ruis, A. R., Haines, K., ... Jung, H. S. (2018). Using epistemic network analysis to identify targets for educational interventions in trauma team communication. *Surgery*, 163(4), 938–943.
- Suthers, D. D., & Desiato, C. (2012). Exposing chat features through analysis of uptake between contributions. In *45th Hawaii International Conference on System Science* (pp. 3368–3377). IEEE.

## Acknowledgments

This work was funded in part by the National Science Foundation (DRL-1661036, DRL-1713110), the Wisconsin Alumni Research Foundation, and the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin–Madison. The opinions, findings, and conclusions do not reflect the views of the funding agencies, cooperating institutions, or other individuals.

# Idea Thread Mapper: Designs for Sustaining Student-Driven Knowledge Building Across Classrooms

Jianwei Zhang and Mei-Hwa Chen  
jzhang1@albany.edu, mchen@albany.edu  
University at Albany, State University of New York

**Abstract:** This research project aims to address the intra- and inter-community challenges associated with student-driven knowledge building over longer terms and across classrooms, in which students take on collective responsibility for charting the course of sustained inquiry and dynamic interaction. We designed the Idea Thread Mapper (ITM) system to support knowledge building in each classroom community as well as idea interaction across communities. At a local classroom level, students engage in focused inquiry and knowledge-building discourse. As their inquiry proceeds, they carry out metacognitive conversations to frame/reframe their collective inquiry directions and processes and synthesize their journey of thinking over time. For cross-community interaction, they selectively share productive idea threads and syntheses in a cross-community space for idea contact with broader knowledge builders. Supported by ITM, we conducted design-based research in a network of Grades 3-6 classrooms. Findings and implications are discussed.

## Introduction

Research on computer-supported collaborative learning (CSCL) has made substantial progress in understanding the sociocultural and cognitive processes of collaborative discourse and knowledge building. However, as a field, we still face the challenge of how to enable sustained educational change and transformation in line with the ethos of CSCL (Chan, 2011; Stahl & Hesse, 2009; Wise & Schwarz, 2017). Tackling this grand challenge requires new research-based innovations to extend CSCL to longer timescales and larger social scales (Stahl, 2013), transcending the traditional boundaries associated with curriculum areas and school years. Aligned with these directions, this paper presents our research to create new learning infrastructures and designs for supporting student-driven knowledge building over long terms across classroom communities.

Design efforts to support student-driven, long-term knowledge building in each classroom need to address two competing demands: to provide the needed classroom guidance and, at the same time, to foster students' high-level agency over creative knowledge processes (Zhang et al., 2018). Extensive research has been conducted to scaffold collaborative learning using collaboration scripts, which are pre-designed to specify, sequence, and distribute various actions and procedures among learners in order to guide effective interactions (Kirschner & Erkens, 2013). Scripted collaboration tends to focus on short-term inquiry activities, in which students are given limited responsibility for charting the process of collaboration and inquiry (cf. Wise & Schwarz, 2017). Long-term, creative knowledge building involves highly dynamic interactions that are difficult to predict and pre-script (Sawyer, 2007). In productive knowledge building communities, students need to take on collective responsibility for charting the course of sustained discourse and inquiry (Scardamalia, 2002). The inquiry process continually evolves and deepens through members' interactive input (Zhang et al., 2011). Deeper problems and knowledge goals are identified over time as new understandings are developed, driving sustained cycles of inquiry. To guide effective and broad classroom implementation of dynamic knowledge building, new research is needed to clarify how such student-driven, open-ended, interactive processes can become socially organized and pedagogically supported while addressing curriculum expectations and contextual constraints (Zhang et al., 2018).

Efforts to support long-term knowledge building in each classroom can further be supported by a larger social infrastructure that allows productive ideas to flow between communities. Such knowledge infrastructures are essential to the productivity of real-world knowledge communities that work as interconnected fields (Csikszentmihalyi, 1999). It is an unsolved challenge for CSCL research to design such accumulative infrastructures and make each community's knowledge progress accessible to other classroom communities. Existing online collaboration systems can easily create a permanent record of student discourse; nevertheless, the knowledge advances achieved in the discourse are not visible or easily interpretable to outsiders who were not part of the discussion. In practice, online discussion spaces are typically reset each academic year for the new student cohort, dumping out the ideas of the previous students. There are emerging efforts to share online discussion spaces between different communities for collaborative knowledge building (e.g. Laferriere, Law, & Montané, 2012). The dominant strategy is to have the different communities directly view each other's online

discussion posts and participate in the same discussion space. However, students often find it difficult to read other communities' online posts and understand their progress and contexts.

The design and research work presented herein aims to address the challenges of how student-driven, long-term knowledge building can be socially organized and how cross-classroom interaction can be supported. Drawing upon our previous research, we designed the Idea Thread Mapper (ITM) system, which was used to conduct design-based research in a network of Grades 3-6 classrooms.

## The design of Idea Thread Mapper

Zhang and colleagues have conducted a series of studies to examine how productive knowledge building communities operate for student-driven, continual idea improvement (e.g. Zhang et al., 2007; Zhang & Messina, 2010; Tao et al., 2015). The findings unveiled a socio-epistemic mechanism by which the student-driven, dynamic inquiry process is organized and scaffolded: emergent reflective structuration. Different from pre-scripted guidance structures, teachers and students engage in ongoing reflection on their inquiry and discourse and co-construct various inquiry structures and resources to support their collective work. The co-constructed structures include shared frames and representations about what the community should investigate, how their inquiry and discourse should be conducted, and who will work with whom in what areas (Tao & Zhang, 2018; Zhang et al., 2018). Guided by the co-constructed structures, students monitor their ongoing knowledge building practices and develop coordinated efforts to advance their collective agenda of inquiry without merely relying on the teacher to lead them through the whole process. Based on the insights gained, we designed a new collaboration system: Idea Thread Mapper (ITM, <http://idea-thread.net>), which interoperates with Knowledge Forum (Scardamalia & Bereiter, 2006) and potentially other collaborative learning platforms. ITM was originally created in 2012 (Chen, Zhang, & Lee, 2013; Zhang et al., 2012) and substantially redesigned and upgraded during 2017-2018. The new ITM incorporates systematic support for co-organizing long-term knowledge building in each classroom community as well as idea interaction across communities. At a local classroom level, students engage in focused inquiry and discourse in their own classroom's discourse space, carry out metacognitive conversations to reflect on shared focuses and insights emerged from the distributed discourse, and synthesize their journey of inquiry including the progress achieved and deeper problems to be addressed. For cross-community interaction, they selectively share productive idea threads and syntheses in a cross-community space for dynamic idea contact with their "buddy classrooms." Across both social levels, students take on high-level responsibility for structuring their unfolding inquiry directions and processes, reviewing progress, and planning for productive collaboration with the support of embedded analytics. Below we summarize the design principles and features.

### Principle 1: Reflective structuration: Co-organize the journey of inquiry as it unfolds

Current collaborative online environments in the forms of online forums and chatting lack support for students to structure and monitor their unfolding, collective journey of inquiry. Students' online discourse tends to focus on teacher-assigned topics and tasks. Also, with students' diverse idea input recorded in distributed online postings, it is often difficult for students and their teacher to monitor the collective knowledge progress achieved in the online discourse (Hewitt, 2001; Suthers, Vatrappu, Medina, Joseph, & Dwyer, 2008; Zhang, 2009). Therefore, to better support students' reflective monitoring and co-organizing of their collective inquiry, ITM incorporates meta-layer representations of student-constructed inquiry structures on the top of the distributed online discourse. The meta-layer inquiry structures serve to capture students' shared directions of inquiry emerging from their interactive discourse, visualize the unfolding strands of inquiry, and document shared progress in each strand of inquiry to inform students' deepening work. The specific features include:

(a) Co-organizing shared wondering areas and inquiry threads: As a high-level structure to guide each knowledge building inquiry that may last over several months, students co-organize their shared wondering areas and specific threads of inquiry (Figure 1) as their work proceeds. Each wondering area is a major direction of inquiry (e.g., brain), which is identified by the classroom members based on their interests and questions. Under each wondering area, members develop one or more inquiry threads, each of which investigates a more specific problem or challenge (e.g. How does the brain work to control body movement?) related to one (or more) wondering area. The wondering areas and threads of inquiry are emergent, co-formed structures, which are similar to "desire lines" —as apposed to pre-planned paths—formed in public spaces (Zhang et al., 2018). In a knowledge building initiative, students begin with open exploration and discourse to develop initial ideas and questions. Through monitoring emerging inquiry interests and evolving needs, they identify high-potential wondering areas and set up idea thread focus, which guide their collaborative discourse and joint inquiry to advance their understandings while continually identifying deeper problems. Each student can select one or more wondering areas as his/her primary focus of inquiry, and adjust his/her focus as the inquiry unfolds.



Students with shared interests form into spontaneous flexible groups to conduct collaborative inquiry and advance their understandings.

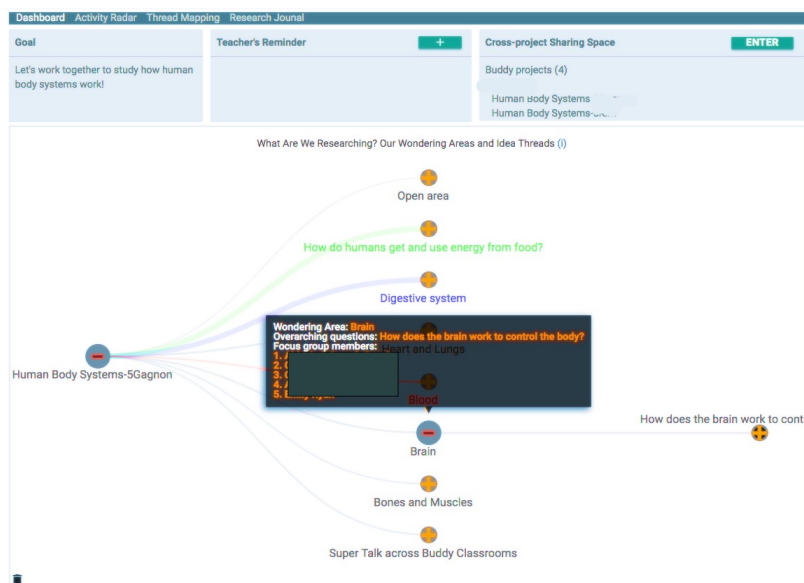


Figure 1. Project dashboard with a visual organizer of the collective wondering areas and idea threads. The thickness of each branch shows the intensity of online contributions in each area.

(b) Visualizing the conceptual, temporal and social profiles of each idea thread: Drawing upon the temporal inquiry-thread analysis developed in our prior work (Zhang et al., 2007), ITM organizes and visualizes students' online discourse in each idea thread as theme-based, temporal interactions. Each idea thread represents a conceptual line of inquiry that involves a sequence of discourse entries--possibly involving several build-on trees--investigating a shared inquiry object over a time period. Figure 2 shows the discourse in an idea thread developed by a group of fifth-graders focusing on how the brain works. The discourse entries (notes) and build-on responses are displayed in a two-dimension space based on the timeline of note creation and the authors involved, showing the temporal and social profile of student contributions in this line of work.

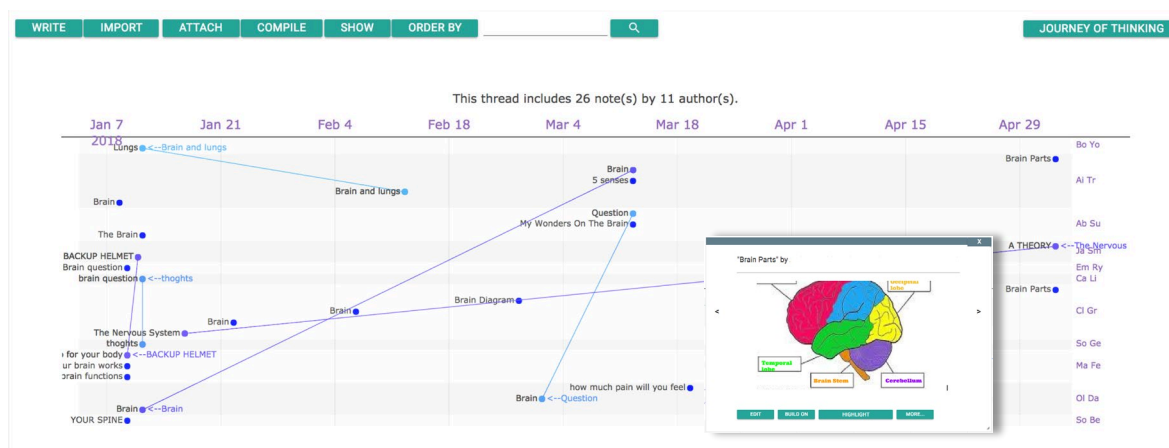


Figure 2. An idea thread developed by a Grade 5 classroom focusing on how the brain works. Each dot represents a note (click to open), and a line connecting two notes shows a build-on connection. The notes are displayed based on the time of creation (x-axis) and author (y-axis).

(c) Reflective documentation of Journey of Thinking in each thread of inquiry: As the conversation in each thread of inquiry proceeds giving rise to new insights as well as deeper issues, the contributors co-compose a Journey of Thinking (JoT) synthesis to deliberate their idea progress. JoT is a reflective "super note" that

includes three sections: problems/issues explored, “big ideas” learned so far, and deeper research needed. Scaffolds are provided in each section to guide student reflection. For example, for the synthesis of the “big ideas” learned, students frame their ideas using “We used to think...now we understand...” (see Table 1).

Table 1: A Journey of Thinking written by a group of 5th graders in the idea thread about breathing

|                         |  |
|-------------------------|--|
| Questions explored:     | [We want to understand:] how do the tissues use the oxygen? How and why do we breathe?   |
| Big ideas learned:      | [We used to think:] That there was not a big process to this.<br>[We now understand:] We get oxygen from air we breath in. The oxygen goes through the thin walls of the air sacs. With the help of hemoglobin, oxygen passes through the air sacs into the blood vessels throughout the body. As that happens CO2 goes out of your body.<br>... |
| Deeper research needed: | [We need to do more learning] about how air get around the body.<br>[We need better theories about] Where is hemoglobin exactly located in the body and does it do anything else besides carry oxygen to the blood, and bring CO2 back into the lungs.<br>[We need to further understand] ...more about tissues.                                 |

(d) Mapping collective progress and connections across different threads of inquiry: At a higher level, the collective landscape of a whole knowledge building initiative is mapped out as clusters of idea threads that investigate interrelated issues through the contributions of all the members (Figure 3). The map of idea threads shows how the different threads of inquiry have evolved over time and further visualizes cross-thread connections, which show student efforts to build on and integrate ideas from the related areas to develop coherent understandings (see analysis in Zhang et al., 2018). The connections include (a) build-on links among the notes in the different idea threads; and (b) connective “bridging contributions,” each of which simultaneously investigates two or more topics focusing on the interrelated issues and cross-cutting ideas. For example, in Figure 3, there is a rich set of notes bridging thread #1 (heart and lungs), #3 (brain control), and #4 (blood), as marked by the dashed vertical lines. Thus, with the visualizations and analytics embedded in the map of idea threads, students can reflect on the temporal process of their inquiry as well as their social participation (e.g. the number of notes and authors involved) and connectedness.

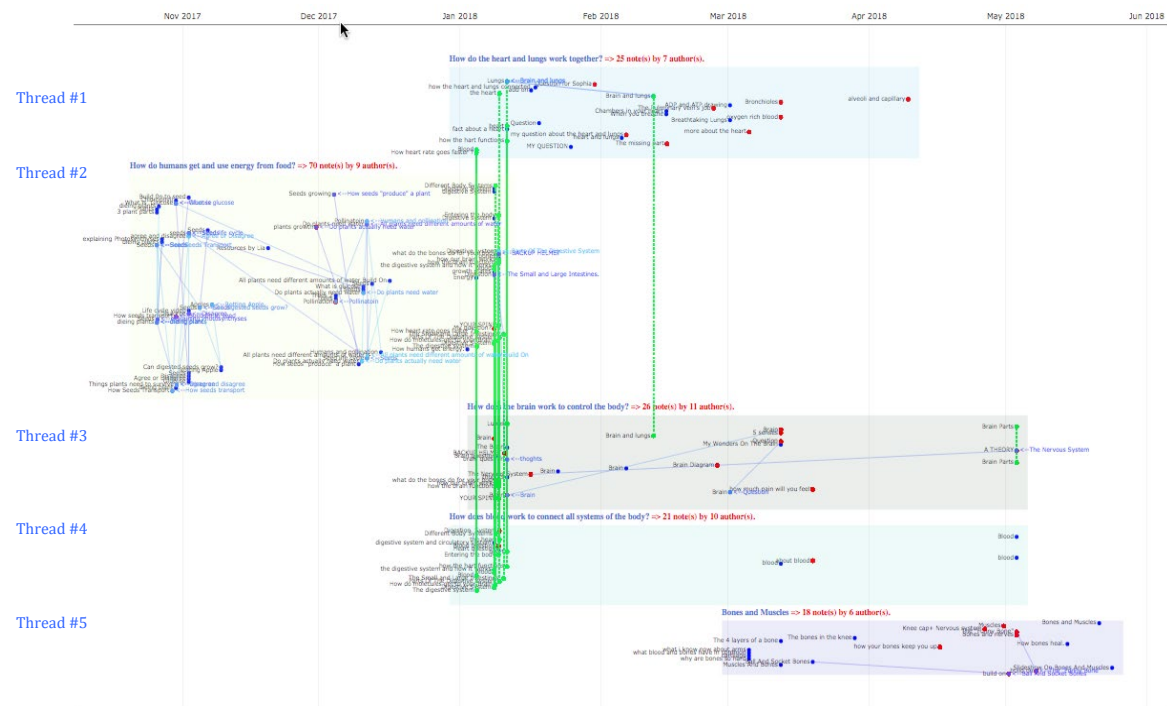


Figure 3. Mapping student discourse in the different idea threads developed in a Grade 5 classroom that studied the human body. Each color stripe represents an idea thread addressing a principal problem, spanning from the

earliest to the latest note created. Each dot shows a note. A line connecting two notes indicates a build-on connection. Dashed vertical lines mark the notes belonging to multiple threads bridging the different lines of discussions.

## Principle 2: Boundary crossing between knowledge building communities

On the basis of the student-driven inquiry in each classroom community, ITM further incorporates supports for cross-community interaction, which is approached using a multi-level, boundary-crossing design (Zhang et al., 2017). As students engage in focused inquiry and conversations and make progress in their own classroom, they reflect on their progress of inquiry and generate JoT super notes, as synthetic knowledge artifacts. These artifacts are used as boundary objects (Star & Griesemer, 1989) to support higher-level interaction across classroom communities.

In the related literature, boundary objects are defined as artifacts (e.g. reports, tools, models) that can be used to bridge the boundaries (discontinuities) between different social worlds (Star & Griesemer, 1989). Objects from a community often have contextual meanings that are not easily accessible to other communities. What makes boundary objects effective for bridging different communities of practice is their interpretative flexibility as a “means of translation” (Star & Griesemer, 1989): they have a structure that is common enough to make them recognizable across the different social contexts and allow different communities to interact and work together. Through interacting with shared boundary objects, members from different communities can identify, understand, and reflect on their different practices, so they can develop better understandings and practices within their own community and potentially create new forms of practices at the intersection of the different communities (Akkerman & Bakker, 2011).

As noted previously, individual online discourse entries are hard to be used as boundary objects for cross-classroom sharing and interaction. Therefore, ITM uses high-level structural representations in the form of idea threads and JoT syntheses as boundary objects. These synthetic artifacts capture the journeys of inquiry taking place in different classrooms using a set of common scaffolds (see Table 1), so they can help the different communities to understand one another’s inquiry directions and knowledge progress. Specifically, ITM incorporates a cross-classroom sharing space for each knowledge building initiative (See Figure 1). The teacher can create “buddy connections” with other classrooms that are studying the same or related content areas, which may come from the same or a different school site. Students can then interact with peers in the buddy classrooms in the following ways:

(a) Ongoing access to the buddy classrooms’ wondering areas and idea threads: Students can see what the buddy classrooms are researching based on their wondering areas and idea threads (see Figure 1), and compare the wondering questions pursued by the different communities. For deeper information, they can also directly view the online discourse in any of the idea threads (in a read-only mode).

(b) Ongoing sharing of the JoT syntheses: Students can access the syntheses generated by the buddy classrooms, and find the most relevant syntheses based on the keyword-based index or using the search tool. As noted previously (Table 1), each JoT synthesizes the progress in a thread of inquiry including the questions explored, “big ideas” learned, and deeper research needed. Thus, by viewing other classrooms’ JoT syntheses, students can get a sense of their journey of inquiry and learn from their “big ideas” and deepening questions. For example, two upper elementary classrooms collaborated in their inquiry about human body systems. While both classrooms investigated how the heart works as part of their inquiry, one classroom’s inquiry questions focused more how the healthy heart functions while the other classroom researched problems about how holes in the heart affect its function. Through reading the heart-related JoT of the other classroom, students realized the different questions asked and discussed the two perspectives to develop coherent understandings (Zhang et al., 2017).

(c) Live Super Talk across buddy classrooms: Students can propose challenging issues as potential topics for cross-classroom discussion, which is called “Super Talk.” For example, in our research, four Grade 5 classrooms studying the human body worked as buddy classrooms. As students made progress in their own classroom understanding how the different body systems work, a few students in one of the classrooms were intrigued by the question of how the human body grows. They proposed this challenging issue as a Super Talk topic shared by the four classrooms. A total of 20 students from the four classrooms contributed 23 notes, looking into how people grow as related to the brain, spine, bones, muscles, skin, digestion, cells, and genetics.

## Principle 3: Embedded analytics to leverage student reflection and idea interaction

ITM integrates a set of analytics and visualizations in the local discourse space and cross-community space (see elements of analytics in Figures 1-3). Specific analytics embedded in the online discourse spaces aim to foster students’ reflective monitoring of the temporal, social, and conceptual profiles of their inquiry. A summary of

the key analytical information is also reported in the Activity Radar, with which students can trace their personal and collective contributions to the various wondering areas and idea threads in a selected time period. Additionally, in collaboration with analytics researchers we tested a set of automated tools to help students monitor inquiry directions, personal contributions, and collaborative connections. For example, to help students monitor emerging topics/areas of inquiry and review progress, ITM provides an automated tool for topical modeling. Specifically, it uses an augmented Latent Dirichlet Allocation (LDA) (Blei, 2012) method to retrieve conceptual topics from student online discourse in relation to topical structures detected from the relevant expert texts (e.g. Wikipedia article). For each topic, the tool further recommends the most relevant online posts, ranked based on semantic relevancy. The user can then review the posts as potential entries for setting up a new idea thread, and define the topic name based on the top keywords. To support student review of each idea thread, ITM further incorporates automated analysis enabled by LightSIDE (formerly known as TagHelper — see Rosé et al., 2008). LightSIDE learns from human-coded discourse data to classify student online posts based on contributions types, including questioning, referencing sources, theorizing, and using evidence (see Tao & Zhang, 2018 for the coding scheme). The automated classification is used to help students monitor the various online discourse moves in each thread of inquiry in order to make informed improvement. To facilitate cross-community interaction, we further tested analytics to gauge the semantic similarity between the idea threads of different communities. Students can find relevant inquiry work conducted by other classrooms, with which they may develop buddy connection and collaboration.

### **Design-based research**

ITM has been tested and refined through our multi-year design-based research conducted in a set of Grades 3-6 classrooms (Zhang et al, 2013, 2014, 2017, 2018; Zhang & Yuan, 2018). Students in each classroom investigated a science topic over multiple months with the support of ITM and Knowledge Forum. Their work integrated whole class knowledge building conversations, individual and group reading, experiments and observations, and so forth. Major questions, ideas, and findings generated through face-to-face activities were contributed online for continual discourse. With their teacher's support and using ITM, students engaged in metacognitive meetings to reflect on the emerging interests and ideas, form shared wondering areas and inquiry directions, and organize their collaborative roles to guide their work in the shared inquiry areas. The visualizations of the wondering areas and collaborative roles were used to guide their monitoring and planning of the ongoing participation and discourse. As the inquiry deepened within their own classroom community, students reviewed productive threads of ideas and composed JoT syntheses, which were shared with their buddy classrooms to stimulate cross-community interaction. Multiple sources of data were collected and analyzed including video records of metacognitive meetings in the classrooms, online discourse and JoT syntheses, individual reflective portfolios, and interviews with the teachers and students. We summarize the main findings below focusing on the within-community and cross-community processes.

### **Co-constructing shared inquiry structures serves to inform students' ongoing discourse and enhance student understandings**

The data analysis documented the processes by which students worked with their teacher to co-construct shared inquiry structures to guide their collective work, including shared frames of collective wondering areas and idea thread topics, mapping of cross-thread connections, and ongoing syntheses of JoT. Such structures served to inform students' ongoing contribution and interaction. With ITM reflection, students were able to conduct more productive and connected discourse: there was a higher proportion of notes identifying deeper questions, using evidence to examine explanations, and integrating different ideas to address challenging issues. Their discourse also became more connected with more build-on links and more connective notes that investigated interconnected inquiry themes and issues (Zhang et al., 2013; 2014; 2018). The enhancement to student inquiry and discourse contributed to improving students' knowledge outcomes. Students with ITM achieved more sophisticated understandings of the science topics as measured based on the content analysis of their reflective portfolios (Zhang et al., 2018). Each student also developed clearer awareness of their community's knowledge goals and progress beyond their personal interests, conducive to enhancing personal learning and collaborative connections (Zhang et al., 2014).

### **Cross-classroom interaction leads to mutual learning and advancement**

We further tested cross-classroom interaction for knowledge building (Zhang et al., 2017, 2018). For example, a design-based research study was conducted in two grades 5/6 classrooms that studied human body systems over a 10-week period (Zhang et al., 2017). As the students conducted focused inquiry and discourse within their

own community, they reviewed productive threads of ideas and posted JoT syntheses in a cross-community space. The syntheses were written by students using ITM's scaffolds for JoT reflection: *Our problem, We used to think...and now we understand..., We need deeper research.* A set of syntheses from the previous classrooms studying human body systems was also posted in the cross-community space. The analyses of the classroom discussions and student interviews suggest that the students engaged in active and substantial interactions with the JoT syntheses from other classrooms. They identified relevant and interesting inquiry topics from other classrooms and compared the different perspectives and ideas, which triggered students' deep reflection on their own inquiry. Further iterations of this design-based research expanded the cross-classroom interaction to more classrooms studying human body systems (Zhang & Yuan, 2018). Content analysis showed that the students' JoT syntheses had a high quality of reflection in capturing solid knowledge progress and deep questions. Social network analysis indicated that through reading the JoT syntheses from their buddy classrooms, students built extensive peer connections focusing on their own inquiry areas as well as the broader scientific concepts. They further identified challenging topics (e.g. how people grow) at the intersection of their interests to conduct cross-classroom collaboration and discourse, putting together their knowledge about the specific body parts to develop coherent explanations.

## Conclusions and implications

ITM and the related studies contribute new strategies to address the intra- and inter-community challenges of long-term knowledge building. Particularly, this research elaborates a reflective structuration approach to implementing long-term inquiry and knowledge building for educational transformation. In a knowledge building initiative (unit) that may extend over multiple weeks or months, the teacher can work with his/her students to co-structure their collective journey of inquiry over time without extensive pre-scripting of the inquiry directions and tasks. Through ITM-supported reflective processes, students can work with their teacher to co-structure the high-level issues about their inquiry, such as what to investigate, through what processes, by whom/with whom as their inquiry unfolds (Zhang et al., 2018; Tao & Zhang, 2018). ITM provides meta-layer representations and tools to capture the co-constructed directions and structures of inquiry, which are visualized to aid members' reflective monitoring and participation. The ITM-based research further elaborates a boundary crossing approach to cross-community interaction. While students engage in focused inquiry and discourse within their own classroom, they generate reflective Journey of Thinking syntheses, as boundary-crossing objects, and participate in the Super Talk to investigate challenging problems together. In sum, the ITM-based technology designs and classroom processes shed light on new designs of socio-technological infrastructures for enabling networks of knowledge building communities, each of which engage in continual idea advancement while building on the knowledge work of other communities. ITM's generic data exchange model supports data exchange with other online discourse environments besides Knowledge Forum. Building on the existing advances, we are currently designing analytics and feedback tools to help teachers and their students monitor emergent inquiry directions and progress in their classroom while discovering potential knowledge connection with other communities.

## References

- Akkerman, S. F., & Bakker, A. (2011). Boundary crossing and boundary objects. *Review of Educational Research, 81* (2), 132-169.
- Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM, 55*(4), 77-84.
- Chan, C. K. K. (2011). Bridging research and practice: Implementing and sustaining knowledge building in Hong Kong classrooms. *International Journal of Computer-Supported Collaborative Learning, 6*, 147-186.
- Chen, M.-H., Zhang, J. & Lee, J. (2013). Making collective progress visible for sustained knowledge building. In N. Rummel, M., Kapur, M. Nathan, & S. Puntambekar (Eds.), *To See the World and a Grain of Sand: Learning across Levels of Space, Time, and Scale: CSCL 2013 Conference Proceedings Volume 1* (pp.81-88). International Society of the Learning Sciences.
- Csikszentmihalyi, M. (1999). Implications of a systems perspective for the study of creativity. In R. J. Sternberg (Ed.), *Handbook of creativity* (pp. 313-335). Cambridge, UK: Cambridge University Press.
- Hewitt, J. (2001). Beyond threaded discourse. *International Journal of Educational Telecommunications, 7*(3), 207-221.
- Kirschner, P. A., & Erkens, G. (2013). Toward a framework for CSCL research. *Educational Psychologist, 48* (1), 1-8.
- Laferriere, T., Law, N., & Montané, M. (2012). An international knowledge building network for sustainable curriculum and pedagogical innovation. *International Education Studies, 5*, 148-160.

- Rosé, C., Wang, Y.-C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., & Fischer, F. (2008). Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in CSCL. *International Journal of Computer-Supported Collaborative Learning*, 3(3), 237-271.
- Sawyer, R. K. (2007). *Group genius: The creative power of collaboration*. New York: Basic Books.
- Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Ed.), *Liberal education in a knowledge society* (pp. 67-98). Chicago, IL: Open Court.
- Scardamalia, M., & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 97-115). New York: Cambridge University Press.
- Stahl, G. (2013). Learning across levels. *International Journal of Computer-Supported Collaborative Learning*, 8(1), 1-12.
- Stahl, G., & Hesse, F. (2009). Classical dialogs in CSCL. *International Journal of Computer-Supported Learning*, 4, 233-237.
- Star, S. L., & Griesemer, J. R. (1989). Institutional ecology, “translations” and boundary objects: Amateurs and professionals in Berkeley’s Museum of Vertebrate Zoology, 1907–39. *Social Studies of Science*, 19, 387–420.
- Suthers, D., Vatrupu, R., Medina, R., Joseph, S., & Dwyer, N. (2008). Beyond threaded discussion: Representational guidance in asynchronous collaborative learning environments. *Computers and Education*, 50, 1103-1127.
- Tao, D., & Zhang, J. (2018). Forming shared inquiry structures to support knowledge building in a Grade 5 community. *Instructional Science*, 46(4), 563-592.
- Wise, A. F., & Schwarz, B. B. (2017). Visions of CSCL: Eight provocations for the future of the field. *International Journal of Computer-Supported Collaborative Learning*, 12(4), 423–467.
- Zhang, J. (2009). Comments on Greenhow, Robelia, and Hughes: Toward a creative social Web for learners and teachers. *Educational Researcher*, 38, 274-279.
- Zhang, J., & Messina, R. (2010). Collaborative productivity as self-sustaining processes in a Grade 4 knowledge building community. In K. Gomez, J. Radinsky, & L. Lyons (Eds.), *Proceedings of the 9th International Conference of the Learning Sciences* (pp. 49-56). Chicago, IL: International Society of the Learning Sciences.
- Zhang, J., & Yuan, G. (2018). Cross-Classroom Interaction for Knowledge Building: A Design Experiment in Four Grade 5 Science Classrooms. Paper presented at the 13th International Conference of the Learning Sciences (ICLS). London, UK: International Society of the Learning Sciences.
- Zhang, J., Bogouslavsky, M., & Yuan, G. (2017). Cross-community interaction for knowledge building in two Grade 5/6 classrooms. In Smith, B. K., Borge, M., Mercier, E., and Lim, K. Y. (Eds.), *Making a Difference: Prioritizing Equity and Access in CSCL, 12th International Conference on Computer Supported Collaborative Learning (CSCL) 2017 (Vol. 1)*. Philadelphia, PA: International Society of the Learning Sciences.
- Zhang, J., Chen, M. -C., Li, H., Zhao, Y., Chen, J., Rajbhandari, B. L., ... & Naqvi, S. (August, 2012). Making collective progress visible: The design and application of Idea Thread Mapper for sustained knowledge building. Knowledge Building Summer Institute, Toronto, ON, Canada.
- Zhang, J., Hong, H.-Y., Scardamalia, M., Teo, C., & Morley, E. (2011). Sustaining knowledge building as a principle-based innovation at an elementary school. *Journal of the Learning Sciences*, 20, 262–307.
- Zhang, J., Lee, J., & Chen, J. (2014). Deepening inquiry about human body systems through computer-supported metadiscourse. Paper presented at the Annual Meeting of American Educational Research Association, Philadelphia, PA.
- Zhang, J., Scardamalia, M., Lamon, M., Messina, R., & Reeve, R. (2007). Socio-cognitive dynamics of knowledge building in the work of nine- and ten-year-olds. *Educational Technology Research and Development*, 55, 117–145.
- Zhang, J., Tao, D., Chen, M.-H., Sun, Y., Judson, D., & Naqvi, S. (2018). Co-organizing the collective journey of inquiry with Idea Thread Mapper. *Journal of the Learning Sciences*, 27, 390-430.

## Acknowledgements

This research was sponsored by the National Science Foundation (#1122573, 1441479). We express our thanks to Drs. Feng Chen, Carolyn Rosé and Marlene Scardamalia for their intellectual input; to our collaborating teachers and their students for their creative classroom work enabling this research; and to the members of our project team who have contributed to the software development and classroom research.

# Location, Location, Location: The Effects of *Place* in Place-Based Simulations

Amanda L. Siebert-Evenstone, University of Wisconsin-Madison, [alevenstone@wisc.edu](mailto:alevenstone@wisc.edu)  
David Williamson Shaffer, University of Wisconsin-Madison, Aalborg University, Copenhagen,  
[david.shaffer@wisc.edu](mailto:david.shaffer@wisc.edu)

**Abstract:** Educators who develop and use place-based curricula argue that engaging students in authentic problem-based learning situated in students' own local place helps them understand the relevance of their academic work, which improves learning. However, while place-based curricula are localized, they are also authentic, problem-based experiences, which have been shown to be effective in their own right. Without controlling for the impact of authenticity, it is not clear whether localization itself improves student learning. In this study, we used a place-based simulation to engage students with authentic problem-based learning about a specific place: a city that is local for some students and not for others. We assessed differences in cognitive, affective, and behavioral outcomes between these two groups. Our results suggest that learning outcomes are better when students engage in a place-based simulation about their own locality, indicating that place does play a critical role in the success of place-based education.

## Introduction

According to theories of *situated learning* (Brown, Collins, & Duguid, 1989), all learning occurs within a specific context. However, learning scientists have also argued that we should also consider how learning is situated, not just that it is situated (Greeno & Middle School Mathematics Through Applications Project Group, 1998). One application of situated learning is *place-based education* in which students learn by engaging in authentic problem-based activities about issues in their local community. When creating place-based curriculum, educators identify content and create localized activities using particular place attributes including the geography, history, social issues, environmental issues, and other specifics of that place. One reason educators may use this method is that research has shown that when students engage in place-based curriculum they show higher increases in knowledge, interest, and pro-environmental behaviors when compared to students in a traditional curriculum. However, despite the centrality of place in environmental education, no studies have tested the effect of place itself, potentially because no existing method can separate the effect of place from the other beneficial pedagogical components associated with this practice.

One way to conduct a controlled study is through the use of simulation. In what follows, we use a *place-based simulation* to reproduce authentic problem-based activities about a real place. We evaluate the effect of place by comparing results between overlapping and non-overlapping physical and virtual locations to identify the effects of localizing curricula. Our results show that localized experiences improve student learning, and such a study could only be conducted using a place-based simulation that enables isolation of effects due to localization.

## Theory

### Place-based education

There are several different versions of place-based education—see, e.g., Gruenewald's (2008) *critical pedagogy of place* and Smith and Sobel's (2010) *place- and community-based education*. However different these approaches are in their specifics, theories of place-based learning generally share two key characteristics: (1) students engage in authentic problem-solving activities; and (2) the problems are situated in the students' own local place—that is, the place where the students themselves live (Gruenewald, 2003; Semken, 2005; Smith & Sobel, 2010; Woodhouse & Knapp, 2000). Semken (2005), for example, describes a place-based curriculum in which students in Arizona use topographic maps to guide the collection of rock, fossil, and soil samples from near their school in order to understand how tectonic forces shaped the geology and resources of their community.

Place-based educators argue that curricula that situate learning in the context of students' own communities help students understand the real-world relevance of their academic work. Unlike more traditional academic approaches which focus on abstract principles, isolated facts, and otherwise decontextualized knowledge, place-based theorists argue that pedagogies that focus on studying issues where students actually live lets them work on concrete, real-world problems in a context they know and care about (Semken et al., 2009).

While in principle place-based education can be developed and applied in any domain, it is most often applied in the context of environmental education. Proponents of place-based environmental education argue that situating learning in students' own local place increases civic and community engagement, interest in environmental issues, and science learning (Powers, 2004; Sobel, 2004). For example, Lieberman and Hoody (1998) conducted a 3-year study of 40 schools across the country, comparing schools using the *Environment as an Integrating Context* curriculum with traditional decontextualized programs. They found that instructional activities based on the local environment around a school increased school achievement across subjects, and increased student engagement and enthusiasm. While these studies did not control for confounding variables, they nevertheless suggest that place-based education can improve teachers' pedagogy and increase student achievement, engagement, and motivation. Based on such studies, Gruenewald (2003) and others (e.g., Smith & Sobel, 2010) claim that place-based learning is the most effective form of environmental education.

## Mechanisms of learning in place-based education

Theorists of place-based learning propose two mechanisms through which place-based learning occurs. One is that place-based curricula offer students multiple options for *authentic experiences* that mirror components of the real-world. Barab and colleagues (2009) suggest that when students work in a setting where they can make consequential decisions, they are more engaged in the learning activities because they believe their actions and decisions matter. Place-based curricula can offer students such experiences a real-world context and a real-world problem to solve, each of which can provide personal and community value. In other words, this argument suggests place-based curricula are effective because they provide consequential and realistic problem-solving activities (Smith and Sobel, 2010). Research in Computer Supported Collaborative Learning more generally has shown authentic experiences (Järvelä, Häkkinen, Arvaja, & Leinonen, 2004) and problem-based activities are beneficial for learning (Hmelo-Silver, 2004). Arastoopour and colleagues (2014) have demonstrated that authentic problem-based activities lead to greater gains in confidence and commitment than students engaging in traditional instruction, and that this method was particularly effective for women.

A second key mechanism of place-based learning is that students are familiar with the context in which their actions are taking place. The first step in many place-based theories calls for *localization* of the curriculum, which refocuses learning to be about the particular place attributes of nearby or regional locations (Woodhouse & Knapp, 2000). By choosing places that students know about, localization provides the opportunity for students to connect to past experiences and envision future experiences in the setting (Smith & Sobel, 2010). Therefore, place-based education provides both a particular kind of pedagogy—authentic, problem-based learning experiences—and a particular kind of setting—the students' own local context. In other words, place-based education is a *pedagogy of place* that combines authenticity and localization.

## Defining local place

While the concept of localization is a key component of place-based education, theorists have struggled to define what constitutes students' own local place and to demonstrate that working in their own local place contributes significantly to students' learning.

Place-based educators emphasize the importance of a local place and often specify features to incorporate into the curriculum from that local place, but rarely identify what constitutes a local place. For example, Semken (2005) defines places as “spatial localities given meaning by human experiences in them” while Gruenewald (2003) and Casey (1997) argue that a local place is one that people live in and know directly. In each of these definitions, nearness is a core component, yet the bounds of this physical distance are never defined. In their seminal book on place-based education, Smith and Sobel (2010) summarize five of the most popular definitions of place-based education, including their own, and outline seven other antecedents to this pedagogy yet never define or specify what constitutes a local place or community. In response to Gruenewald and others, Stevenson (2008) questions the concepts of local and place, asking what educators and theorists actually mean when they advocate for learning about a local place.

There is a body of research, primarily in sociology and urban planning, on *attachment* to a place: that is, the extent to which people feel that they belong in or to a place, and their perception that the place is familiar and “their own” (Devine-Wright, 2013). More than 75% of place attachment studies focus on the neighborhood scale (Lewicka, 2011), which Devine-Wright (2013) and others (Feitelson, 1991) argue limits understanding of the different scales of place and that place attachments may occur at multiple scales. To address this issue, Laczko (2005) collected data from 24 countries and found that after their high affinity to their country, United States citizens felt closest to their state community rather than their neighborhood or city.

## Testing a pedagogy of place



Although place-based education may sound promising and beneficial, a fundamental assumption of this pedagogy has not been tested: the effect of the *place* itself. As described above, extant research comparing place-based education with traditional curricula has not controlled for the simultaneous effects of *authentic experiences* and *localization*, both of which are hypothesized as key components of a pedagogy of place. When compared with traditional classrooms, do students in place-based curricula do better because of localization, or because activities are authentic, or merely that they leave the classroom? Or is there some synergy between these elements of a pedagogy of place that make it more effective than either localization or authentic experiences alone?

In other words, while place-based curricula are localized, they are also authentic, problem-based learning experiences, which have been shown to be effective in their own right. Without controlling for the impact of authenticity it is not clear whether localization improves student outcomes.

One way to disambiguate the impact of localization from the other features of place-based learning would be to compare the effects of local and non-local experiences of the same place-based curriculum. Such a study would require a curriculum that could present students with authentic, problem-based issues situated in an *existing* place, but in one that is not the student's *own* local place. So, for example, students would have to use Semken's (2005) place-based geology curriculum (described above) based on the specific rock formations near the Colorado Plateau even though they did not live in Arizona. However, the curriculum has students use local topographic maps to identify places to collect rock, fossil, and soil samples to learn about local rocks and local ground and surface water resources. To adapt this place-based curriculum to another part of the country would require finding a different local land formation for collecting samples that had similar petrologic and hydrologic properties – or rewriting the curriculum to focus on different geologic issues. In other words, changing the location of the curriculum potentially requires a substantial reworking of the original curriculum.

An alternative way to address this issue would be to simulate the two main components of place-based education: *authentic problem-based learning* and *specific local contexts*.

Simulations are technological environments that are designed to reproduce events, places, experiments, and processes from the real world (Dawley & Dede, 2014). In such environments, designers can replicate problem-solving contexts, real-world activities, professional tools, common social interactions, and realistic work products. For example, Chesler and colleagues (2015) created *Nephrotex* where students can role-play as engineering interns at a biotechnology company designing filtration membranes for kidney dialysis. Simulations such as *Nephrotex* offer students the ability to engage in *authentic problem-based learning* about core disciplinary ideas in realistic settings.

## Current study

In this study, we explore the impact of a *place-based simulation* (PBS): a simulation that engages students with authentic problem-based learning *about a real location*. One example of a PBS is the virtual internship, *Land Science*, in which students assume the role of interns at a fictitious urban planning firm. Students engage in realistic professional work in a meaningful real-world context (Chesler et al., 2015) by performing the kinds of tasks that urban planners do in their training: they receive materials that urban planners use, such as research reports, ecological impact surveys, and communications from concerned citizens, which provide information about revenue, water pollution, housing, and other issues. Students engage in these authentic problem-based activities to develop and justify land-use plans that meet the needs of competing stakeholders. Additionally, *Land Science* simulates these activities in the specific context of an urban planner working in Lowell, MA, using content, history, stakeholders, environmental indicators, and maps that are place-specific to that city. Through participation in *Land Science*, students learn about complex eco-social systems (Bagley & Shaffer, 2011) and learn to think like urban planners in the context of a real city (Beckett & Shaffer, 2005). In this sense, *Land Science* is a curriculum that is *based on a real place* and replicates core elements of place-based pedagogy.

PBSs like *Land Science* have several advantages. Typically, in order to adapt a place-based curricula to a new location, educators would need to adapt existing materials to ensure that the curriculum is both specific to a place and uses authentic practices. Because PBSs are based on a real place and use authentic activities, students can interact with the same curriculum from anywhere. As a result, the experience could be local for some students and not for others depending on the geographic location of the student. In other words, we can use *Land Science* to hold authenticity constant and vary the localization of the curriculum. We can, therefore, test differences in student performance between students who might consider Lowell, MA their own *local* place against students in other areas across the country who are engaging in a *non-local* experience.

Of course, in order to test the effect of place, meaningful measures for environmental education must be used to identify differences. Commonly, environmental educators are interested in cognitive, affective, and behavioral changes. For example, the North American Association of Environmental Education (2013) identified the goal of environmental education as creating environmentally literate students, where literacy is contingent

upon changing skills, values, and behaviors instead of solely increasing content knowledge. Therefore, we assessed differences between local and non-local experience using common cognitive, affective, and behavioral outcomes. To address cognitive changes, we assessed students' ability to identify an example of a scientific model—a model that resembles, represents, and/or summarizes the functionality of an object or phenomenon by making a particular feature of the world easier to understand, define, quantify, visualize, or simulate (Lehrer & Schauble, 2006)—because modeling is a critical practice of science in general and of environmental science in particular (Bagley & Shaffer, 2011). Because the virtual internship addresses learning about urban planning, we assessed interest in cities and the environment as an affective outcome. Finally, we assessed changes in future behavior by adapting a civic engagement measure about knowledge and ability to engage in community problems (Flanagan, Syvertsen, & Stout, 2007) that was adapted for middle and high school students and included school as a possible community.

In this study, our primary research questions examine the relationship between online place and location of play across three different outcomes. In this study, we ask:

1. Do students who engaged in a local PBS have *higher ability to identify a science model* after the simulation than students who engaged in a non-local PBS?
2. Do students who engaged in a local PBS have *higher changes in interest* than students who engaged in a non-local PBS?
3. Do students who engaged in a local PBS have *higher changes in civic engagement* than students who engaged in a non-local PBS?

## Methods

### Land Science virtual internship

In the virtual internship *Land Science*, students explore the environmental and socio-economic impacts of land-use decisions. Students role-play as urban planning interns at Regional Design Associates, an urban planning firm developing a land-use plan for the city of Lowell, Massachusetts. Students work individually and in teams to develop a rezoning plan for Lowell that addresses the demands of various community groups who advocate for environmental and socio-economic issues such as wildlife protection, job creation, housing, and controlling air and water pollution. Students use iPlan, a geographic information system model, to evaluate the impacts of land use. Each zoning plan cannot address all of the stakeholder's concerns, so students must make and justify decisions about which demands to meet and how to meet them. The virtual internship takes ten to fifteen hours to complete.

### Local and non-local simulations

Following Laczko (2005) as discussed above, we categorized students who engaged in the *Land Science* PBS about the state they live in and know about as students who had a *local* experience of a place-based curriculum. We categorized students who engaged in the *Land Science* PBS about a state where they do not live as students who had a *non-local* experience of a place-based curriculum. Because simulations can be played from many sites about endless simulated locations, the *Land Science* PBS provided a platform to test the differences between local and non-local experiences of a place-based curriculum.

### Participants

We collected pre- and posttest responses from 94 middle and high school students who participated in one of 10 different implementations of *Land Science* (6 local and 4 non-local). All students participated in *Land Science* as part of an informal science learning experience run by a teacher in a place-based science education center. We examined data from 68 students who engaged in a local PBS and 26 students who engaged in a non-local PBS. To be included in analyses, students must have completed the internship and answered all three sets of questions.

### Measures

Students completed an online survey before (pretest) and after (posttest) participating in *Land Science*. The survey had three components: (1) *Scientific Modeling*, (2) *Interest*, and (3) *Civic Engagement*.

#### Cognitive measure: Scientific modeling score

The survey asked students to write an example of a scientific model. Answers were scored with 1 point if they provided an accurate example of a scientific model and 0 points otherwise. For example, some students incorrectly identified elements of the scientific method, such as “in a scientific model, you need to do basic research and to have a hypothesis.” While correct responses would include accurate examples of a model such as a “globe” or

“double helix” model of DNA. Two raters achieved acceptable interrater reliability on a subset of the data ( $n = 126$ , Cohen’s  $\kappa = 0.91$ , Shaffer’s  $\rho(0.65) < 0.01$ ), and a single rater coded the rest of the data (please see Shaffer (2017) for more details).

#### Affective measure: Interest score

The survey asked students to respond to an open-ended question about their interest in cities and the environment. Table 1 lists the five codes that identify varying types of interest. Two raters coded a subset of 120 responses. We compared their rate of agreement and all five scores had acceptable agreements (Cohen’s  $\kappa < 0.65$ ,  $\rho(0.65) < 0.01$ ) resulting in a weighted Kappa of 0.72. A single rater scored the rest of the data.

Table 1: Categories to code student interest in the cities and/or the environment

| Category  | Points | Example   |
|---|--------|---|
| No Interest   | 0      | “I have no interest in cities or the environment”   |
| Interest in <i>either</i> a specific city or cities <i>or</i> the a specific environmental issue          | 25     | “I’m interested to learn about how cities work”   |
| Interest in <i>both</i> a specific city or cities <i>and</i> the a specific environmental issue or issues | 50     | “I want to learn about plans that help out cities and the environment.”   |
| Interest in the <i>relationship</i> between cities and environment generally                              | 75     | “I’m interested in learning how we affect the environment and what we can do to reduce to effect our cities have on the environment.” |
| Interest in a <i>specific relationship or relationships</i> between cities and the environment            | 100    | “I’m interested in making room for the entire population of the city, without overcrowding, while keeping the environment intact.”    |

#### Behavioral measure: Civic engagement score

The survey asked students to self-report their knowledge and ability to engage in school and community problems by answering 11 four-point Likert Scale questions adapted from Flanagan, Syversten, and Stout (2007), questions such as: “I know ways of addressing community problems” and “I would be able to find and examine research related to the issue.” We summed the student scores for each of the 11 questions and rescaled the result to create a *Civic Engagement Score* from 0 to 100 for each student at two time points (pretest and posttest), where a score of zero indicated strong *disagreement* on all questions and 100 indicated strong *agreement* on all questions.

### Analyses

We constructed a series of nested multiple regression analyses to predict the change in outcome for students’ *Interest* and *Civic Engagement Scores*. Because *Scientific Modeling* was a dichotomous variable we constructed a series of nested logistic regressions to predict this outcome. In each analysis, we tested the following predictors: location, pretest Civic Engagement Score, pretest Interest Score, and pretest Scientific Modeling Score. Pretest measures were included in each set to control for differences in students’ academic ability levels across the 10 implementations included in the study. We computed the Bayesian Information Criterion (BIC) to determine which of the models had the best balance between model fit and model simplicity. For each set of models, we chose the model with the lowest BIC and reported the associated tests for that model.

### Results

#### Change in scientific modeling score

The optimal nested logistic regression for Scientific Modeling included location and scientific modeling pretest as predictors which had a significant regression equation,  $\chi^2(2) = 21.25$ ,  $p < 0.001$  (see Table 2). Students who were able to identify a scientific modeling example at the beginning of the game had a 3.09 times higher relative chance of identifying a model after the internship than students who were unable to identify models at the beginning of the internship. For students in Massachusetts, the relative chance of identifying a model is approximately 2.34 times higher than students who played *Land Science* in a different state. There was no effect of civic engagement pretest and interest pretest.

Table 2: Logistic regression results predicting change in Scientific Modeling Score

| Model | Intercept | Predictors |           |            | Model Comparison |        |
|-------|-----------|------------|-----------|------------|------------------|--------|
|       |           | Modeling   | Location: | Civic Egmt | Interest         | Pseudo |

|      | Pretest                      | MA                            | Pretest                      | Pretest                    | R <sup>2</sup>              |                |
|------|------------------------------|-------------------------------|------------------------------|----------------------------|-----------------------------|----------------|
| Null | 0.22<br>(0.14) <i>1.25</i>   |                               |                              |                            | --                          | 295.35         |
| I    | -0.24<br>(0.19) <i>0.79</i>  | 1.07***<br>(0.29) <i>2.90</i> |                              |                            | 0.05                        | 286.72         |
| IIA  | -0.01<br>(0.17) <i>0.99</i>  |                               | 0.75*<br>(0.31) <i>2.12</i>  |                            | 0.02                        | 294.55         |
| IIB  | -0.36<br>(0.60) <i>0.70</i>  |                               |                              | 0.01<br>(0.01) <i>1.01</i> | 0.00                        | 299.72         |
| IIC  | 0.29<br>(0.29) <i>1.33</i>   |                               |                              |                            | -0.00<br>(0.00) <i>1.00</i> | 0.00<br>300.63 |
| III  | -0.53*<br>(0.22) <i>0.59</i> | 1.13***<br>(0.30) <i>3.09</i> | 0.85**<br>(0.32) <i>2.34</i> |                            |                             | 0.07<br>284.80 |
| IV   | -1.71*<br>(0.76) <i>0.18</i> | 1.21***<br>(0.31) <i>3.34</i> | 0.90**<br>(0.33) <i>2.46</i> | 0.02<br>(0.01) <i>1.02</i> | 0.00<br>(0.01) <i>1.00</i>  | 0.08<br>292.55 |

Standard errors are reported in parentheses. Odds ratios are reported in italics. \* $p < 0.05$ , \*\* $p < 0.01$ , and \*\*\* $p < 0.001$ .  $n = 94$ .

### Change in interest score

The optimal nested regression model for Interest Score based on BIC included location and interest as predictors which had a significant regression equation ( $F(2,91) = 13.26, p < 0.001$ ), with an adjusted  $R^2$  of 0.21 (see Table 3). Participant change in interest scores decreased by 0.51 percent for each 1-point increase in the pretest score indicating that students with lower interest at the start of the game had higher increases in interest. On average, students in Massachusetts increased their interest score by 21.41 percentage points more than the scores for students in a different state increased. Because we scaled these tests to have a maximum value of 100, this difference in scores between local and non-local students, controlling for their pretest score, would be the equivalent of increasing two letter grades. There was no effect of civic engagement pretest or example pretest.

Table 3: Regression results predicting change in Interest Score

| Model | Predictors         |                    |                   |                    |                  | Model Comparison        |        |
|-------|--------------------|--------------------|-------------------|--------------------|------------------|-------------------------|--------|
|       | Intercept          | Interest Pretest   | Location: MA      | Civic Egmt Pretest | Modeling Pretest | Adjusted R <sup>2</sup> | BIC    |
| I     | 28.13***<br>(5.52) | -0.49***<br>(0.13) |                   |                    |                  | 0.12                    | 914.34 |
| IIA   | -3.85<br>(5.93)    |                    | 20.39**<br>(6.97) |                    |                  | 0.08                    | 919.29 |
| IIB   | 19.18<br>(13.07)   |                    |                   | -0.12<br>(0.19)    |                  | -0.01                   | 927.21 |
| IIC   | 9.88*<br>(4.82)    |                    |                   |                    | -0.02<br>(0.07)  | -0.01                   | 927.56 |
| III   | 13.21<br>(6.91)    | -0.51***<br>(0.13) | 21.41**<br>(6.46) |                    |                  | 0.21                    | 908.15 |
| IV    | 6.79<br>(13.73)    | -0.53***<br>(0.13) | 22.58**<br>(6.70) | 0.12<br>(0.18)     | 0.03<br>(0.06)   | 0.20                    | 916.66 |

Standard errors are reported in parentheses. \* $p < 0.05$ , \*\* $p < 0.01$ , and \*\*\* $p < 0.001$ .  $n = 94$ .

### Change in civic engagement score

The optimal nested regression model for Civic Engagement Score based on BIC included location and civic engagement pretest score as predictors which had a significant regression equation ( $F(2,91) = 15.83, p < 0.001$ ), with an adjusted  $R^2$  of 0.24 (see Table 4). Participant civic engagement scores decreased by 0.34 percent for each 1-point increase in the pretest score indicating that students with lower pretest scores showed greater change in civic engagement score. On average, students in Massachusetts had a change in civic engagement score that was 9.74 percent higher than the change in score for students playing *Land Science* in a different state. Because we scaled these tests to have a maximum value of 100, the difference in post-test scores between local and non-local students, controlling for their pretest score, is the equivalent change of one full letter grade. There was no effect of example pretest or interest pretest.

Table 4: Regression results predicting change in Civic Engagement Score

| Model | Predictors |                    |              |                  |                  | Model Comparison        |        |
|-------|------------|--------------------|--------------|------------------|------------------|-------------------------|--------|
|       | Intercept  | Civic Egmt Pretest | Location: MA | Modeling Pretest | Interest Pretest | Adjusted R <sup>2</sup> | BIC    |
| I     | 32.86***   | -0.39***           |              |                  |                  | 0.17                    | 776.96 |

|     |                    |                    |                    |                |                |       |        |
|-----|--------------------|--------------------|--------------------|----------------|----------------|-------|--------|
|     | (5.88)             | (0.09)             |                    |                |                |       |        |
| IIA | -1.63<br>(2.88)    |                    | 12.10***<br>(3.39) |                |                | 0.11  | 783.61 |
| IIB | 10.22***<br>(2.35) |                    |                    | 0.06<br>(0.03) |                | 0.02  | 792.62 |
| IIC | 7.04*<br>(2.94)    |                    |                    |                | 0.00<br>(0.07) | -0.01 | 795.82 |
| III | 22.74***<br>(6.53) | -0.34***<br>(0.08) | 9.74**<br>(3.18)   |                |                | 0.24  | 772.30 |
| IV  | 17.92*<br>(7.00)   | -0.32***<br>(0.09) | 10.36**<br>(3.20)  | 4.89<br>(2.87) | 0.03<br>(0.06) | 0.25  | 777.93 |

Standard errors are reported in parentheses. \* $p < 0.05$ , \*\* $p < 0.01$ , and \*\*\* $p < 0.001$ .  $n = 94$ .

## Discussion

Our results suggest that curricula situated in a specific, real place are more effective when that place is local to the learners. To predict change in each outcome, we included the location, relevant pretest, as well as the other pretests to see which variables might explain differences in the score. For each of the three outcome variables, however, the nested regressions showed that only the relevant pretest and location were significant predictors of the change in outcome. Models that included both the relevant pretest and location best explained the change in outcome and that this model was more parsimonious than models that included all four predictors. These results suggest that even though all students engaged in the same authentic activities and problem-based learning about a real place, we found that students showed different changes in outcomes based on their location. When students engaged in a simulated place-based curriculum about their state, they had higher changes in civic engagement, interest, and ability to identify a scientific model than students in other states.

These results have implications both for the study of place-based pedagogies and for the use of simulations in education more generally.

While the importance of local place has been assumed in place-based education research, this is the first work we are aware of that provides empirical evidence to support this claim. It may be that previous studies have made this assumption because of the physical constraints that limit the places students can experience in an engaging and immersive way. Our results suggest that when students experience a simulation of authentic and problem-based learning about a specific place, it makes a difference if it is their own local place. Thus, as Gruenewald (2003) and other place-based theorists argue, environmental education may be more effective when it engages students in the places where they live.

More broadly, though, researchers who study educational simulations argue that one benefit of these tools is that educators can create curricula that can be accessed from locations anywhere in the country or world (Dawley & Dede, 2014). This means that students are not limited in their access to sound pedagogies based on the resources of their own communities. However, this study suggests that local context has a powerful influence on how students perceive and experience an online curriculum. There are surely important ideas that are best understood through the lens of distant or imagined worlds, but it may be that students are more swayed by Thoreau's (1981) argument that "it is not worth the while to go round the world to count the cats in Zanzibar." As a result, teachers may need the ability to customize one-size-fits-all simulations—or perhaps more accurately, one place fits all simulations—to use them more effectively in environmental science classrooms.

This study has several limitations. First, it describes a small number of students in a small number of locations where students could not be randomly assigned to a locality. However, this reestablishes the problems in measuring the effect of place. While we were not able to use randomization and the sample size was not large enough to account for differences in the implementations such as length or background work by teachers, we would be quick to point out that we have no reason to believe that those are systematically related to the place in which they were implemented. As a result, although we are currently working to expand this analysis to a larger sample size with more locations, the implications of this study should be generalized with caution. Second, this study tested only one definition of local place and does not explore all of the complex issues involved with conceptions of space or place. Future analyses could examine the impact of different spatial scales for locality to better understand the relationship between localization and learning. Finally, this study focused on three limited outcomes measured in a pretest-posttest design. As a result, it provides no information on the mechanisms by which locality impacted learning in this environmental education simulation, and future studies are needed to better understand the effect of locality on science content and practice learning within place-based simulations.

Despite these limitations, these results suggest that students have higher changes in civic engagement, interest, and ability to identify a scientific model when they engage in a place-based simulation about their own local place—and that advocates for place-based education are correct in arguing that situating curriculum in students local place has a positive impact on student outcomes. Place does matter.

## References

- Arastoopour, G., Chesler, N., & Shaffer, D. (2014). Epistemic persistence: A simulation-based approach to increasing participation of women in engineering. *J. of Women and Minorities in Sci. and Eng.*, 20(3), 211–234.
- Bagley, E. A. S., & Shaffer, D. W. (2011). Promoting civic thinking through epistemic game play. In R. Ferdig (Ed.), *Discoveries in gaming and computer-mediated simulations: New interdisciplinary applications* (pp. 111–127). Hershey, PA: IGI Global.
- Barab, S. A., Gresalfi, M., & Arici, A. (2009). Why educators should care about games. *Ed. Leadership*, 67(1), 76–80.
- Beckett, K. L., & Shaffer, D. W. (2005). Augmented By Reality: the Pedagogical Praxis of Urban Planning As a Pathway To Ecological Thinking. *Journal of Educational Computing Research*, 33(1), 31–52.
- Brown, J., Collins, A., & Duguid, P. (1989). Situated Learning and the Culture of Learning. *Ed. Research*, 18(1), 32–42.
- Casey, E. S. (1997). *The Fate of Place: A Philosophical History*. Berkeley, CA: University of California Press.
- Chesler, N. C., Ruis, A. R., Collier, W., Swiecki, Z., Arastoopour, G., & Shaffer, D. W. (2015). A novel paradigm for engineering education: Virtual internships with individualized mentoring and assessment of engineering thinking. *Journal of Biomechanical Engineering*, 137(2), 1–8.
- Dawley, L., & Dede, C. (2014). Situated learning in virtual worlds and immersive simulations. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of research on educational communications and technology* (4th ed., pp. 723–734). New York.
- Devine-Wright, P. (2013). Think global, act local? The relevance of place attachments and place identities in a climate changed world. *Global Environmental Change*, 23(1), 61–69.
- Feitelson, E. (1991). Sharing the globe: The role of attachment to place. *Global Env. Change*, 1, 396–406.
- Flanagan, C. A., Syvertsen, A. K., & Stout, M. D. (2007). Civic Measurement Models: Tapping Adolescents' Civic Engagement. *The Center for Information & Research on Civic Learning & Engagement*, May(May), 40.
- Greeno, J. G., & Middle School Mathematics Through Applications Project Group. (1998). The Situativity of Knowing, Learning, and Research. *American Psychologist January 1998*, 53(1), 5–26.
- Gruenewald, D. A. (2003). Foundations of place: A multidisciplinary framework for place-conscious education. *American Educational Research Journal*, 40(3), 619–654.
- Gruenewald, D. A. (2008). The best of both worlds: a critical pedagogy of place. *Env. Ed. Research*, 14(3), 308–324.
- Hmelo-Silver, C.E. (2004). Problem-Based Learning: What and How Do Students Learn? *Educational Psychology Review*, 16(3), 235–266.
- Järvelä, S., Häkkinen, P., Arvaja, M., & Leinonen, P. (2004). Instructional Support in CSCL. In J. W. Strijbos, P. A. Kirschner, & R. L. Martens (Eds.), *What We Know about CSCL* (pp. 115–119). Kluwer Academic Publishers.
- Laczko, L. S. (2005). National and Local Attachments in a Changing World System: Evidence from an International Survey. *International Review of Sociology*, 15(3), 517–528.
- Lehrer, R., & Schauble, L. (2006). Cultivating model-based reasoning in science education. In K. Sawyer (Ed.), *Cambridge Handbook of the Learning Sciences* (pp. 371–388). Cambridge, MA: Cambridge University Press.
- Lewicka, M. (2011). Place attachment: How far have we come in the last 40 years? *J. of Env. Psych.*, 31(3), 207–230.
- Lieberman, G. A., & Hoody, L. L. (1998). *Closing the achievement gap: Using the environment as an integrating context for learning. Results of a nationwide study*. Poway, CA: Science Wizards.
- NAAEE. (2013). North American Association for Environmental Education, <http://www.naaee.net/>.
- Powers, A. L. (2004). An evaluation of four place-based education programs. *Journal of Env. Education*, 35(4), 17–32.
- Semken, S. (2005). Sense of Place and Place-Based Introductory Geoscience Teaching for American Indian and Alaska Native Undergraduates. *Journal of Geoscience Education*, 53, 149–157.
- Semken, S., Freeman, C., Watts, N., Neakrase, J., Dial, R., & Baker, D. (2009). Factors that influence sense of place as a learning outcome and assessment measure of place-based geoscience teaching. *Elect. J. of Sci. Ed.*, 13(2), 136–159.
- Shaffer, D. W. (2017). *Quantitative Ethnography*. Madison, WI: Cathcart Press.
- Smith, G. A., & Sobel, D. (2010). *Place- and community-based education in schools*. New York, NY: Routledge.
- Sobel, D. (2004). Place-based Education: Connecting Classroom and Community. *The Orion Society*.
- Stevenson, R. (2008). A critical pedagogy of place and the critical place(s) of pedagogy. *Env. Ed. Res.*, 14(3), 353–360.
- Thoreau, H. D. (1981). *Walden and other writings*. Bantam Classics.
- Woodhouse, J. L., & Knapp, C. E. (2000). Place-Based Curriculum and Instruction: Outdoor and Environmental Education Approaches. *ERIC Clearinghouse on Rural Education and Small Schools*, 1–8.

## Acknowledgments

This work was funded in part by the National Science Foundation (DRL-1661036, DRL-1713110), the Wisconsin Alumni Research Foundation, and the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison. The opinions, findings, and conclusions do not reflect the views of the funding agencies, cooperating institutions, or other individuals.

# The Role of Funds of Knowledge in Online Search and Brokering

Wendy Roldan, Paola Vanegas, Laura Pina, Carmen Gonzalez, and Jason Yip  
wr4@uw.edu, pvdt@uw.edu, lpina@uw.edu, cmgonzal@uw.edu, jyip@uw.edu  
University of Washington

**Abstract:** Lower-socioeconomic status, immigrant parents who are English language learners often work collaboratively with their children to search the internet. Family members rely on each other's language and digital literacy skills in this collaborative information problem solving process known as online search and brokering (OSB). While previous work has identified the ecological factors that impact OSB, research has not yet distilled the specific learning processes behind such collaborations. From a case study analysis of three families, this work explores the funds of knowledge that children and parents rely on as they engage in collaborative learning experiences through OSB. We demonstrate how in-home computer supported collaborative processes are often informal, collaborative, social, and highly relevant to solving real-life information challenges. Our work shows how parents and children draw on their funds of knowledge when they search collaboratively, with and for their family members, to build their collective knowledge of technology and problem-solving.

## Introduction

Approximately 8 million U.S. children have at least one immigrant parent who is an English language learner (Zong, Batalova, & Hallock, 2018). Lower-socioeconomic status immigrant parents often rely on their children's language and digital literacy skills to address family needs (Eksner & Orellana, 2012). In these families, adults and children work with each other to search the internet, in a collaborative information problem-solving process we call *online search and brokering* (OSB) (Pina et al., 2018; Yip, Gonzalez, & Katz, 2016). For these families, children's responsibilities extend beyond traditional chores to holding the role as the family's primary problem-solver for critical family needs through online information searches (Pina et al., 2018). While previous work has identified online search and brokering practices (Yip et al., 2016) and the ecological factors that influence online search and brokering (Pina et al., 2018), research has not yet distilled the specific learning processes behind such family collaborations. While online information problem-solving models (Brand-Gruwel, Wopereis, & Walraven, 2009) contribute to our understanding of digital learning and information processing, they do not fully explain the sociocultural, collaborative learning processes that occur as families engage in online search and brokering.

Our research focuses on Latin American families, the fastest growing population in the U.S. (Pew Research Center, 2018). Latino children are projected to make up about a third of the K-12 enrollment by 2023 and with more than half of them living in immigrant families (Foxen & Mather, 2016), many of them are searching for critical information online for their English language learning family members. Education researchers engaged in sociocultural historical approaches with children from non-dominant backgrounds explain that learning is an ongoing process that is not divided into separate characteristics of individuals and contexts (Gutiérrez & Rogoff, 2003). Therefore, we believe it is important to understand the learning that occurs in intergenerational, bilingual, information problem-solving family collaborations. Previous research with families whose households have been traditionally viewed as low-resourced, applies a funds of knowledge framework to understand and nurture the strategic knowledge and skills that are often overlooked (Moll, Amanti, Neff, & Gonzalez, 1992). Using a funds of knowledge framework (Moll et al., 1992) this work explores the learning processes that occur in Latino families, when family members collaboratively search for information online. We examine a case study of three focal Latin American, lower-socioeconomic status, English language learner families as they engaged in collaborative, computer supported, online search and brokering. Our research questions are: **RQ1**. What funds of knowledge do bilingual children and their English language learner parents rely on when attempting to solve information problems using the Internet?; and **RQ2**. What are the collaborative learning processes around online information searches in English language learner families as they work together?

## Background

This work builds on literature in individual and collaborative search processes, the phenomenon of search and brokering, joint media engagement, and sociocultural historical theory on learning as a cultural, ongoing activity. We draw on research that investigates how individuals search for information online to make sense of the learning that happens within individual search processes. However, in applying individual problem-solving models for online search we find gaps with respect to understanding how learning occurs in intergenerational and bilingual collaborative searches. Thus, we turn to the literature on language brokering that focuses on intergenerational and

bilingual problem-solving between youth and their parents to help us fill the gaps. To strengthen our literature review on online search and brokering with families, we find that joint media engagement is a particularly useful framework to examine the learning that is happening as families use media together and in the home.

### **Individual and collaborative search processes**

A rich body of work explores how individuals search for information online. Models like *Information Problem-Solving for the Internet* offer insights into the set of skills individuals need to search for information online. These skills include: defining, searching, scanning, processing, and organizing information (Brand-Gruwel et al., 2009). While we know much about the learning that happens within individual search processes, less is known about the process of how people collaborate together to solve information problems and the learning processes behind it (Stahl, Koschmann, & Suthers, 2006). Research on online searching as a collaborative process between individuals (Morris, 2013), shows the opportunity to reinforce and learn search skills from exposure to others' strategies (Foss et al., 2012). But much of this work has focused on skilled peer adults searching remotely together using online tools (Morris, 2013). Further, research with families engaged in informal collaborative relationships with technology suggests that participation in these activities can nurture knowledge of managing information and promoting self-direction of one's learning (Jenkins, 2006). This strand of research provides insights into the relationship between collaboration and learning but focuses on collaborative engagement in technologically mediated activities that are not information problems. From our literature review, we conclude that research on intergenerational, bilingual, collaborative search processes and learning is limited thus we turn to the communications literature that explores language brokering with youth in the home.

### **Online search and brokering (OSB)**

Recent literature explores how parents and children work collaboratively to address family information needs using digital resources, in a process known as online search and brokering (Pina et al., 2018). Research on immigrant youth shows how language brokering is part of everyday life as children and parents work together to address family needs (Orellana, Dorner, & Pulido, 2003). Bilingual youth work as translators and interpreters for their immigrant parents, which opens the families' access to resources and information on education, health, and finances. As families collaborate to address family needs in everyday language-brokering events, different levels of skills and expertise are leveraged, and knowledge becomes co-constructed (Eksner & Orellana, 2012). As the digital divide narrows, scholars have also explored how children become *technology brokers* to introduce and teach their parents new technologies (Nelissen & Van den Bulck, 2018). There are significant differences in how children help their parents with technology across families from different socio-economic status (Brown, Bakken, Nguyen, & Von Bank, 2007). In high-socioeconomic status families, children drive the adoption of mobile applications, technology for entertainment, and educational purposes. On the other hand, in families from a lower-socioeconomic status, children help connect their adult family members to critical information needs (e.g., finances, health, well-being) (Nelissen & Van den Bulck, 2018). Collaborative family engagement with technology also differs significantly across families. To further contextualize our work as parents and children use technology together, we draw on literature from joint media engagement.

### **Joint media engagement (JME)**

The nuanced interactions that occur as parents and children engage in online search and brokering is a form of joint media engagement that has not been previously studied (Pina et al., 2018). The phenomenon of joint media engagement helps us understand the experiences of people using media together as they view, play, contribute, search, and create with both traditional and digital media (Sobel et al., 2017; Takeuchi & Stevens, 2011). Previous research with families defines joint media engagement as the process of learning between children and parents as they co-create meaningful connections among interests, experiences, and representations using all forms of media and technologies that are present in children's lives (Takeuchi & Stevens, 2011). However, much of the research on joint media engagement emphasizes learning together through gaming, entertainment, and education (Gee, Takeuchi, & Wartella, 2018). Our work builds on prior research with families and technology by identifying the funds of knowledge that children and their parents draw on to problem-solve and co-create meaningful connections that extend beyond play and family fun to solve critical family needs.

### **Theoretical framework – funds of knowledge**

We apply a funds of knowledge framework to examine the collaborative, intergenerational, bilingual, information problem-solving learning processes that occur during online search and brokering with an asset-based perspective. A funds of knowledge framework allows us to understand the historically accumulated and culturally developed bodies of knowledge that are essential for household and individual well-being (Moll et al., 1992). We posit that a funds of knowledge approach allows us to capture the tacit knowledge parents and children draw on when



engaging in online search and brokering, knowledge that is not necessarily taught but critical to family well-being. This framework allows us to push the conversation away from deficit models of learning for non-dominant students towards honoring the learning that occurs in the household and everyday activities of Latino youth and adults as they search and broker for information online collaboratively. Previous research with families whose households have been traditionally viewed as low-resourced uses funds of knowledge to refer to the strategic knowledge and skills that exist within these families (Moll et al., 1992). A person's funds of knowledge can be described as their accumulated life experiences, the skills and knowledge they use to navigate everyday life, and their cultural-historical academic and personal background knowledge. Our work builds on prior funds of knowledge research in learning and education, by identifying the knowledge and skills found in households that members use to solve information problems as a computer supported collaborative learning process.

As parents and children engage in everyday online search and brokering, we explore the learning processes that occur through a sociocultural lens by applying a funds of knowledge framework. Sociocultural scholars propose that the educational movement toward equity will occur, "as we create learning environments that connect in deep ways to the life experiences of all students" (Nasir, Rosebery, Warren, & Lee, 2006). A sociocultural lens provides new insights on issues of race, culture, and learning (Nasir et al., 2006). Learning, as a cultural process of engagement in repertoires of practices, is a process in which individual development is understood in cultural and historical contexts (Gutiérrez & Rogoff, 2003). Building on this body of work, we recognize understanding learning requires a focus on how individuals engage and participate in particular everyday activities and how they draw on artifacts, tools, and others to solve problems (Nasir & Hand, 2006).

## Methods

For this study, we adhered to the standards and practices of a case study examination (Merriam & Tisdell, 2015). We focused on three Latin American lower-socioeconomic status, English language learner families in the Pacific Northwest and their collaborative learning experiences through engagement in online search and brokering.

This exploratory, qualitative study with three focal families is a part of a larger qualitative study. Between July 2016 and June 2017, we visited 23 families in an urban area of the Pacific Northwest, U.S., within a 32-kilometer radius of our research institution. Our participants included parents, grandparents, and children aged 10-17. Most of the adults that were a part of our study were born in Mexico, did not have a college-degree, worked in service industries, and represented a lower-socioeconomic population. At the time of this study, a tense political climate existed for Latino families therefore we relied on community center networks, local cultural events, and a paid community liaison as a part of our recruitment strategies. We conducted two separate in-home family visits per family. In-home visits allowed participants to feel comfortable in their usual search and brokering environment and allowed researchers to observe search practices within the home's digital infrastructure. In our first in-home visit (V1), we conducted separate adult and youth retrospective interviews to contextualize where and how they search. In V1 we gathered family members' perspective on their practices, strategies, and challenges when searching for information online. The interview protocol was adapted from previous work on in-home media studies (Katz & Gonzalez, 2016). Each interview (45-60 mins) was audio recorded and transcribed. Parent interviews were in Spanish and youth interviews in English. For our second in-home visit (V2), we focused on directed internet search tasks between adult-youth dyads. Visit 2 was audio recorded and screen interactions were video recorded. In V2, participants engaged in a set of imposed search tasks prompted by researchers and historical tasks of prior online searches. We take a deep dive into three specific families from this larger data set (Table 1).

Table 1: Demographics of the families

| ID | Relationship (Age)              | Occupation    | Adult birthplace | Grade completed by adult          |
|----|---------------------------------|---------------|------------------|-----------------------------------|
| 2  | Mother (39), Daughter (10)      | Hospitality   | Mexico           | Primary, 4 <sup>th</sup> grade    |
| 17 | Grandfather (62), Grandson (12) | Food Industry | Mexico           | Secondary, 7 <sup>th</sup> grade  |
| 22 | Mother (42), Daughter (15)      | Housekeeping  | Mexico           | Secondary, 10 <sup>th</sup> grade |

Our selection criteria for these three families are the following. (1) Each family offered insights into different household structures that influence the online search and brokering process. One family was a mother-daughter dyad within a single-parent home, another was a grandparent-grandson dyad within a seven-person home, and the third was a mother-daughter dyad within a four-person family home. (2) In a comparative analysis each family allowed us to see how despite the differences across family structures, we observed themes in the funds of knowledge used and the learning processes occurring during online search and brokering. (3) Each family offers insights into different types of devices used in the home to search for information.

For this paper, we analyzed data from V1 and V2 with three families to explore our research questions. To analyze the data, two researchers used open coding with constant comparative analysis (Merriam & Tisdell, 2015). This iterative open-coding was done for both parent and child interviews and for an analysis of V2 video

data. Two researchers open-coded the data independently for themes such as technology and language brokering, information problem-solving, relationships, perceptions of learning, and strategies and challenges to searching. We coded and compared the themes to develop further categories for analysis and then systematically compared and contrasted the themes between researchers. Following open-coding, we used axial coding to make connections between funds of knowledge categories and subcategories. We performed a constant sorting and comparative analysis until theoretical saturation was reached and no new codes were generated. Finally, we looked at these themes and interpreted them using a funds of knowledge analytical lens. Data on search tasks from our second visit allowed us to triangulate interview data from our first visit to provide a rich analysis of our data.

## Findings

For each family, we provide a case description for context and describe a search task in-depth. We then discuss each case using a funds of knowledge analytical lens (Moll et al., 1992) to help us document the learning and knowledge in practice that is a part of online search and brokering as a daily household routine (Gutiérrez, Morales, & Martinez, 2009). We highlight the funds of knowledge that each family member brings into the collaborative information problem search process from their home and community lives (Moll et al., 1992). Each family engages in online search and brokering differently with devices, set-ups, and strategies. All quotes are translated from Spanish. We use [ ] to indicate typing on digital devices. All names used in the cases are pseudonyms, with Name<sup>A</sup> for adult, and Name<sup>C</sup> for child.

### Family 2. Norma<sup>A</sup> and Mia<sup>C</sup>

*Case Description.* Mia<sup>C</sup> is Norma<sup>A</sup>'s 10-year-old youngest daughter who helps her mom with searches on a more regular basis compared to her older sisters. Mia<sup>C</sup> is a developing yet savvy searcher, often struggling with spelling search words but consistently recognizing the need for descriptive terms to narrow her search. At the time of the study, Norma<sup>A</sup> indicated she did not feel confident with technology but was aware of technology resources through her local community networks. In V1, we learned Norma<sup>A</sup> and Mia<sup>C</sup> engaged in collaborative online search and brokering when Norma<sup>A</sup> discovered someone had taken money out of her bank account fraudulently. The woman from the bank taught Mia<sup>C</sup> how to access Norma<sup>A</sup>'s online banking information for her. During V2, we saw this financial search information problem revisited. For one of the tasks, Norma<sup>A</sup> wanted to find the closest bank near her home. Mia<sup>C</sup> clicked the search bar (Figure 1, left), and began to type using the tablet screen keyboard and her two index fingers, [*the cosest bank*]. Mia<sup>C</sup> paused as the search engine suggestions came up for *the closest bank of america* and *the closest bank of america to me*. Mia<sup>C</sup> continued to type into her search [*the cosest bank of america*] and pressed the search button to generate results. Mia<sup>C</sup> scanned the results and clicked on the Google visual map (Figure 2, right). Mia<sup>C</sup> hovered her finger over the options and explained to her mom what the results meant. Mia<sup>C</sup> noted Option A is the closest because it appeared first.

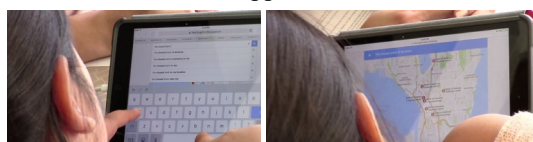


Figure 1. Norma and Mia searching for the closest bank using Google Maps.

### Case analysis. Knowledge of linguistic translational practices

In this case, we notice how Norma<sup>A</sup> and Mia<sup>C</sup> use their funds of knowledge about each other's assets to complete their financial information search task of finding the bank nearest to them. Mia<sup>C</sup> built on her cultural repertoires of linguistic practice as she (a) translated her mom's inquiry about finding the closest bank from Spanish to English; (b) searched using key descriptive terms in English that would narrow her search for banks nearest her; (c) scanned the information presented in English; (d) relied on visuals and maps to find answers; and (e) translated the results back to Norma<sup>A</sup>. After the search was completed, Mia<sup>C</sup> explained she did not understand maps, but Norma<sup>A</sup> used her knowledge of spatial geography and relied on her lived experiences of navigating this area to understand where the bank was. Norma<sup>A</sup> and Mia<sup>C</sup> helped each other process, translate, and increase their collective understanding of finding and reading a map. We see knowledge of linguistic translational practices used in their intergenerational, bilingual, online search and brokering practices. In this example, we see Mia<sup>C</sup> engaging in a range of practices that challenge deficit notions of students' repertoires developed across non-school settings (Gutiérrez et al., 2009). As Mia<sup>C</sup> collaborative searches with and for her mom, she uses linguistic knowledge, problem-solving and search knowledge, and translation knowledge. Each of these funds of knowledge Mia<sup>C</sup> uses to search could become future resources for her learning across settings and practices, inside and outside the classroom (Gutiérrez et al., 2009).

## Family 17. Carmelo<sup>A</sup> and Mateo<sup>C</sup>

*Case Description.* Mateo<sup>C</sup> is Carmelo<sup>A</sup>'s 12-year-old youngest grandson who helps Carmelo<sup>A</sup> with most of his searches. Carmelo<sup>A</sup> owns a restaurant and is a lifelong learner who desires to learn more about technology. Mateo<sup>C</sup> is a visual and audio searcher who feels comfortable with technology and uses google voice to search for information. Similar to Mia<sup>C</sup>, Mateo<sup>C</sup> struggles with spelling and instead prefers to say rather than type a word. In V1, Carmelo<sup>A</sup> told us that he searches the internet to buy things for his restaurant and preferred images to make sense of information online. In V2, we observed Carmelo<sup>A</sup> and Mateo<sup>C</sup> search together for new industrial stoves knobs for Carmelo<sup>A</sup>'s restaurant. Carmelo<sup>A</sup> wanted to buy goods on Amazon for his restaurant. Mateo<sup>C</sup> clicked on the Amazon app on a smartphone and clicked on the search bar (Figure 2, left). He moved his fingers from the search bar to the keyboard. Mateo<sup>C</sup> showed Carmelo<sup>A</sup> how to go to the search bar. Carmelo took the smartphone and typed [*bottons for*] in Spanish. Mateo<sup>C</sup> interrupted and asked if this was for clothing, but Carmelo<sup>A</sup> explained this was for stoves. When Carmelo finished typing, the Amazon app returned images of what Carmelo was looking for. As the discussion continued, there was a disconnect between what Mateo<sup>C</sup> thought they were looking for. Mateo<sup>C</sup> typed word [*butten*] into the Amazon search bar (Figure 2, center). Mateo<sup>C</sup> said the search was not giving him what he wanted. Carmelo<sup>A</sup> took back the phone, left the Amazon app, clicked the Google app, and typed [*how do you spell bottons in English*]. Carmelo<sup>A</sup> said, "There it is! This is how I do it. But it takes me a long time." At this point Carmelo<sup>A</sup> asked Mateo<sup>C</sup> to teach him how to search using voice to be more agile. Mateo<sup>C</sup> clicked on the voice icon on Google search (Figure 2, right) and told him to just talk. Carmelo<sup>A</sup> took the phone up to close to his mouth and said, "I want to look for buttons for the stove," but the search did not work. Mateo<sup>C</sup> explained, "You have to push this button and then go. You push it and go." Carmelo<sup>A</sup> used the voice assistant again, "I want to look for buttons for the stove." Carmelo<sup>A</sup> looked and smiled at the search result images of the knobs he was looking for, having learned a new technology skill (Figure 2, right).

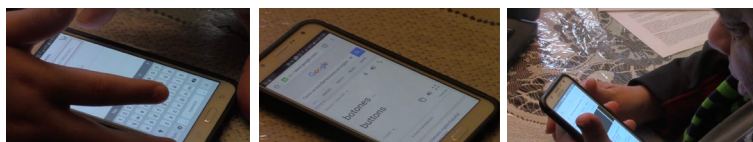


Figure 2. Carmelo and Mateo searching for stove knobs using the Amazon app and Google voice assistant.

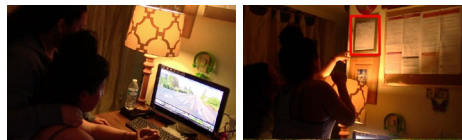
### Case analysis. Knowledge of problem-solving strategies

In this case, we believe Carmelo<sup>A</sup> and Mateo<sup>C</sup> rely on each other's funds of knowledge to explore different search strategies when they face the challenge of not knowing the translation between buttons and knobs in Spanish. Carmelo relies on his lived experiences running the restaurant to explain to Mateo<sup>C</sup> what he was looking for. Mateo<sup>C</sup> sets the search up for Carmelo<sup>A</sup> by (a) previously configuring the Amazon app settings to Spanish; (b) showing him where to enter his search terms; and (c) clarifying what it is they are looking for to make sure the language translation is correct. When they get stuck, Carmelo<sup>A</sup> uses his previous knowledge related to persistence to show Mateo<sup>C</sup> how he uses the Internet to problem-solve translational challenges. Further, this example shows how each other's funds of knowledge are taken up as they explore alternative search strategies moving from the Amazon app, to Google search, to Google search voice assistant. Carmelo<sup>A</sup> observes Mateo<sup>C</sup> using the Google voice assistant and learns how to do it himself through Mateo<sup>C</sup>'s scaffolds. The two help each other conceptualize the problem, explore different internet resources to push away from challenges, and learn new language and technology skills. We see knowledge of alternative strategy exploration being taken up by Carmelo<sup>A</sup> and Mateo<sup>C</sup> as they expand their knowledge of technology to uncover the shared meaning of knobs/buttons in their multisensory (auditory and visual), intergenerational, bilingual, and online search and brokering practices.

## Family 22. Romelia<sup>A</sup> and Amy<sup>C</sup>

*Case Description.* Amy<sup>C</sup> is Romelia<sup>A</sup>'s 15-year-old daughter. Romelia<sup>A</sup> is a stay-at-home mom and relies on Amy<sup>C</sup> to search for information related to the needs of others in the home. Romelia<sup>A</sup> watches videos on YouTube to learn how to do things like braid hair or find new recipes. Amy<sup>C</sup> is a searcher who uses multiple strategies in action as she searches for her mom using her knowledge about viruses, advertisements, and web browsing. When the two engage in collaborative online search and brokering together they often sit side-by-side, co-viewing on the family desktop computer. During our first visit, we learned Romelia<sup>A</sup> wished she knew how to help her husband search for construction jobs. In V2, we observed how Amy<sup>C</sup> and Romelia<sup>A</sup> search for construction jobs for the father. Amy<sup>C</sup> asked, "What jobs do you want to help find for papa?" (Figure 3, left). Romelia<sup>A</sup> responded, "Like companies, like companies that have jobs." Amy<sup>C</sup> typed, [*construction companies hiring near me*] and clicked on the second search result, Monster.com. Romelia<sup>A</sup> noted, "Make sure the search results are local." Amy<sup>C</sup> narrowed the search by selecting the U.S. state and typed [*construction*] in the Monster.com search box. The search results come up and Amy<sup>C</sup> translated the job descriptions from English to Spanish for her mom. As Amy<sup>C</sup>

simultaneously interpreted, translated, and scanned the results, she also explained why she was not clicking on job postings that require a technical or engineering degree. However, Amy<sup>C</sup> noted she was stuck. Romelia<sup>A</sup> suggested adding the word *siding* and Amy typed in [*constructing citing* (sic)], which did not result in what they needed. Amy<sup>C</sup> asked, “*citing*, right?” They stepped away from the computer and look for other resources to refine their search terms. Romelia<sup>A</sup> pointed to the father’s contracting certification that showed how to spell *siding* the correct way (Figure 3, right). Amy typed, [*construction companies hiring siding installers in (U.S. state)*]. She translated the information to her mom. Romelia<sup>A</sup> asked, “How much do those jobs pay?” Amy noted it was about \$20 an hour. Amy<sup>C</sup> asked, “Do you want jobs for him or for his company?” Romelia<sup>A</sup> stated, “I want to look for jobs for a contractor, not jobs that hire by the hour, I want to look for big jobs because he has his company.”



**Figure 3.** Romelia and Amy searching for construction siding jobs using online and physical resources.

### Case analysis. Knowledge of search resources

Romelia<sup>A</sup> and Amy<sup>C</sup> discuss, contribute, and move from online resources to resources in the home to move past challenges. When Amy<sup>C</sup> began the search, Romelia<sup>A</sup> used her geospatial knowledge and lived experiences of finding a job to ask Amy to narrow the search for local jobs. Amy<sup>C</sup> used her knowledge to (a) click on a job search engine; (b) scan the job postings; (c) translate linguistically the search results; and (d) explain her interpretation of the job descriptions. When the two got stuck, Romelia<sup>A</sup> offered new search terms (*siding*) while Amy<sup>C</sup> continued to scan and interpret website search results. Using knowledge of alternative resources available to them beyond the Internet they collectively (a) step away from the computer; (b) find the father’s contracting license; and (c) correct their spelling of their search terms. They build on each other’s funds of knowledge to conceptualize the problem, find alternative resources to spell the search term, and explain what type of job they are looking for based on the father’s qualifications. We argue knowledge of search resources moves fluidly between the two as they translate across language, across physical and online, and across technical linguistic practices.

## Discussion

*Funds of Knowledge in Online Search and Brokering.* In this study, we found that each individual family member contributes their own knowledge to collectively build household knowledge resources. Each child is a part of a larger social structure within the home where they rely and share their knowledge with siblings, parents, and grandparents. Across all families, we observed how Mia<sup>C</sup>, Mateo<sup>C</sup>, and Amy<sup>C</sup>’s role as the primary searcher in their home carried “the trace of prior social relations,” building on their prior actions while co-creating new knowledge with adults (Bakhtin, 1981; Vossoughi, Jackson, Chen, & Roldan, Under Review; Vygotsky, 1978). The adults usually have knowledge for the search conceptualization such as geospatial understanding, spelling clarifications, or knowledge of family resources. The children usually have critical technology proficiency and linguistic translational knowledge to search, scan, process, gather, and translate online information. During collaborative searches, each family member continually builds on the knowledge of each other’s assets. This knowledge informs the ways in which children and adults set up future learning opportunities for each other within online search and brokering practices.

Further, we highlight how family members developed resilience, as they faced structural and systemic challenges to searching for information (language, access to social resources, schooling). Resilience was reinforced through funds of knowledge and resource sharing as family members relied on each other’s knowledge, strategies, and skills when met with a challenge during their search process (Berkes & Ross, 2013). Across all families, we observed resilience at the level of the individual and the household as interrelated, with family members actively developing their shared resilience through capacity building and collaborative learning while engaged in online search and brokering (Berkes & Ross, 2013). The families persisted and relied on their funds of knowledge (translational practices, problem-solving, search resources) collaboratively to solve their information problems when they got stuck or tried to problem-solve through the unknown. For the majority of the adults in this study, the United States is not their home country. With this information, we can see the funds of knowledge related to resilience, problem-solving through new situations, and strategies to move beyond challenging situations while engaging in online search and brokering in our data. Building on sociocultural theories of learning, we offer this analysis as a way to help us construct classrooms that draw on the funds of knowledge that students bring from their home and everyday experiences to promote holistic and equitable development and learning (Gutiérrez & Rogoff, 2003).

*Collaborative Learning Processes in Online Search and Brokering.* In contrast to highly individualized

instructional systems (e.g., formal school), we demonstrate learning processes in-the-home are often informal, collaborative, highly social, and highly relevant to solving real-life information access challenges. Through our thematic analysis of the funds of knowledge that adults and children draw on while they engage in collaborative online search and brokering with their families, we offer a nuanced understanding of computer supported collaborative learning processes happening within the home. We contribute to studies of joint media engagement (Takeuchi & Stevens, 2011) by identifying the funds of knowledge Latino families draw on to problem-solve and co-create solutions that extend beyond play to solve critical family needs. Joint media engagement is not just about learning together through gaming, entertainment, or educational technologies. Instead, we highlight the ways that family members work together as a group to solve information problems using online connected technologies. Work within CSCL notes that group cognition forms as a result of collaborative knowledge building in which meaning is created across the utterances of different people (Stahl, 2006). In our cases, all of the adults and youth act in a joint activity to problem-solve for their family needs. The shared construction of meaning occurs as both adult and child engage through intersubjectivity of language, technological interactions, information problem-solving, and family funds of knowledge. For many of these families, the group remains at the synchronous level, in which adult and child work together around an interactive device (e.g., desktop, tablet, laptop, smartphone) to collaboratively solve their family information needs. We believe our research in this area contributes to CSCL as there is a need to identify how technology and online information problem-solving for family needs supports how people learn, and uncovering what challenges lie ahead.

*Implications for Educators.* As online search and brokering plays a major part of family practices, our work highlights the funds of knowledge that are evident as parents and children search together, which educators can utilize in the classroom to design instructional materials that are relevant to students' family responsibilities. Overall, while our findings focus on a specific group of Latin American families with English language learner parents from a lower-socioeconomic status, family information search and collaboration happen across all contexts and ages. For example, other immigrant English language learner families might search for health information for elderly family members together. Individuals must use similar processes of finding information together, making sense of that information, and translating that information by closely drawing on their funds of knowledge at each step of the information problem-solving model. By examining a) the funds of knowledge families use in online search and brokering and b) the computer supported collaborative learning processes in online information search, we are better able to provide instructional design implications to help schools, libraries, and community centers attend to culture in understanding students' learning. Our work opens up new opportunities to bridge home and school through the computer supported collaborative learning process that are embedded in children's every day practices. Our findings can help educators consider family roles when designing curricula. Future work could examine the unique challenges of English language learning students and how their search practices might fluctuate across formal and informal contexts.

## Conclusion

We argue that an intergenerational online search and brokering process is different compared to collaborative online information problem-solving that happens between classroom peers or between co-workers. Our research shows how both parents and children draw on their funds of knowledge when they search collaboratively, with and for their family members, to build their collective knowledge of technology and problem-solving. Through the metaphor of a jigsaw puzzle, we can begin to understand how collaborative problem-solving takes first, the identification of knowledge and skills and secondly, the learning processes behind figuring out how to can arrange, turn, and shift the puzzle pieces, in this case the funds of knowledge, to solve the information problem at hand. The CSCL community can benefit from an understanding of how online search and brokering is a form of collaborative learning around technologies, given it is the daily reality for millions of bilingual children.

## References

- Bakhtin, M. M. (1981). *The dialogic imagination: Four essays by mm bakhtin* (m. holquist, ed.; c. emerson & m. holquist, trans.). Austin: University of Texas Press.
- Berkes, F., & Ross, H. (2013). Community Resilience: Toward an Integrated Approach. *Society & Natural Resources*, 26(1), 5–20. <https://doi.org/10.1080/08941920.2012.736605>
- Brand-Gruwel, S., Wopereis, I., & Walraven, A. (2009). A descriptive model of information problem solving while using internet. *Computers & Education*, 53(4), 1207–1217.
- Brown, B. B., Bakken, J. P., Nguyen, J., & Von Bank, H. G. (2007). Sharing information about peer relations: Parent and adolescent opinions and behaviors in Hmong and African American families. *New Directions for Child and Adolescent Development*, 2007(116), 67–82.
- Eksner, H. J., & Orellana, M. F. (2012). Shifting in the zone: Latina/o child language brokers and the co-

- construction of knowledge. *Ethos*, 40(2), 196–220.
- Foss, E., Druin, A., Brewer, R., Lo, P., Sanchez, L., Golub, E., & Hutchinson, H. (2012). Children's search roles at home: Implications for designers, researchers, educators, and parents. *Journal of the American Society for Information Science and Technology*, 63(3), 558–573.
- Foxen, P., & Mather, M. (2016). Toward A More Equitable Future: The Trends and Challenges Facing America's Latino Children. Retrieved from <http://publications.nclr.org/handle/123456789/1627>
- Gee, E., Takeuchi, L., & Wartella, E. (Eds.). (2018). *Children and families in the digital age: learning together in a media saturated culture*. New York: Routledge, Taylor & Francis Group.
- Gutiérrez, K. D., Morales, P. Z., & Martinez, D. C. (2009). Re-mediating literacy: Culture, difference, and learning for students from nondominant communities. *Review of Research in Education*, 33, 212–245.
- Gutiérrez, K. D., & Rogoff, B. (2003). Cultural ways of learning: Individual traits or repertoires of practice. *Educational Researcher*, 32(5), 19–25.
- Jenkins, H. (2006). *Confronting the challenges of participatory culture: Media education for the 21st century*. An occasional paper on digital media and learning. John D. and Catherine T. MacArthur Foundation.
- Katz, V. S., & Gonzalez, C. (2016). Toward meaningful connectivity: Using multilevel communication research to reframe digital inequality. *Journal of Communication*, 66(2), 236–249.
- Merriam, S. B., & Tisdell, E. J. (2015). *Qualitative research: A guide to design and implementation*. John Wiley & Sons.
- Moll, L. C., Amanti, C., Neff, D., & Gonzalez, N. (1992). Funds of knowledge for teaching: Using a qualitative approach to connect homes and classrooms. *Theory into Practice*, 31(2), 132–141.
- Morris, M. R. (2013). Collaborative search revisited. *Proceedings of the conference on computer supported cooperative work*. ACM, 2013.
- Nasir, N. S., & Hand, V. M. (2006). Exploring Sociocultural Perspectives on Race, Culture, and Learning. *Review of Educational Research*, 76(4), 449–475.
- Nasir, N. S., Rosebery, A. S., Warren, B., & Lee, C. D. (2006). Learning as a Cultural Process: Achieving Equity Through Diversity. In *The Cambridge handbook of: The learning sciences*. (pp. 489–504). New York, NY, US: Cambridge University Press.
- Nelissen, S., & Van den Bulck, J. (2018). When digital natives instruct digital immigrants: Active guidance of parental media use by children and conflict in the family. *Information, Communication & Society*, 21(3), 375–387.
- Orellana, M. F., Dorner, L., & Pulido, L. (2003). Accessing assets: Immigrant youth's work as family translators or para-phrasers. *Social Problems*, 50(4), 505–524.
- Pew Research Center. (2018). *Hispanic/Latino Demographics*. Retrieved from <http://www.pewresearch.org/topics/hispaniclatino-demography/>
- Pina, L. R., Gonzalez, C., Nieto, C., Roldan, W., Onofre, E., & Yip, J. C. (2018). How Latino Children in the US Engage in Collaborative Online Information Problem Solving with their Families. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW), 140.
- Sobel, K., Bhattacharya, A., Hiniker, A., Lee, J. H., Kientz, J. A., & Yip, J. C. (2017). It wasn't really about the Pokémon: Parents' Perspectives on a Location-Based Mobile Game (pp. 1483–1496). ACM Press.
- Stahl, G. (2006). *Group cognition: Computer support for building collaborative knowledge*. Cambridge, Mass: MIT Press.
- Stahl, G., Koschmann, T. D., & Suthers, D. D. (2006). *Computer-supported collaborative learning*. Takeuchi, L., & Stevens, R. (2011). The new coviewing: Designing for learning through joint media engagement. Presented at the New York, NY: The Joan Ganz Cooney Center at Sesame Workshop.
- Vossoughi, S., Jackson, A., Chen, S., & Roldan, W. (Under Review). Embodied pathways and ethical trails: Studying learning in and through the choreography of social relations.
- Vygotsky, L. (1978). Interaction between learning and development. *Readings on the Development of Children*, 23(3), 34–41.
- Yip, J.C., Gonzalez, C., & Katz, V. (2016). The learning experiences of youth online information brokers. In C. K. Looi, J.L. Polman, U. Cress, & P. Reimann, P. (Eds.). *Transforming Learning, Empowering Learners: The International Conference of the Learning Sciences (ICLS)* (Vol. 1, pp. 362 – 369).
- Zong, J., Batalova, J., & Hallock, J. (2018). *Frequently Requested Statistics on Immigrants and Immigration in the United States*. Migration Policy Institute.

## Acknowledgments

We thank the families that allowed us into their homes and the individuals that contributed to this research.



# Collaboration on a Massive Scale – Conceptual Implications of *the Crowd*

Norma Reichelt, Christine Bussian, Christoph Richter, Heidrun Allert, and Lars-Arne Raffel  
reichelt@paedagogik.uni-kiel.de, bussian@paedagogik.uni-kiel.de, richter@paedagogik.uni-kiel.de,  
allert@paedagogik.uni-kiel.de, lars.raffel@email.uni-kiel.de  
Christian-Albrechts-Universität zu Kiel, Institute of Educational Science

**Abstract:** The question of how to engage large user groups in web-based learning arrangements is consistently raised in CSCL discourse. Recently the interest in crowd phenomena of joint knowledge creation and inquiry as well as their adaptability within institutionalized learning is brought together under the term of mass collaboration. This paper proposes a conceptualization of the crowd since learner involvement, sociality, collaboration and knowledge construction, coordination and regulation efforts, as well as assessment must be fundamentally thought anew, when we take a step away from a small group setting. Thus, we put forward a design framework that provides perspectives on CSCL environments on a massive scale and thereby challenges our understanding of collaborative learning.

## Introduction

In the last years, the masses have played a decisive role in raising social, economic, political and environmental issues and mass movements like Occupy, the Arab Spring or #blacklivesmatter have drastically established unprecedented ways of public discourse and engagement. In science and education, large accumulations of learners, researchers or civic participants are addressed in open education initiatives like Massive Open Online Courses (MOOCs) or Open Educational Resources (OERs) and get involved in open science or citizen science projects. Further, it is also relevant to mention phenomena of joint development and creation (for example open source software development or the global online encyclopedia Wikipedia) as well as various possibilities for distributed financing, fundraising or forms of investigative journalism (for example the investigation platform Bellingcat). With these new forms of participation we notice new forms of collaboration and knowledge practices that are inseparably linked with technological developments especially in the fields of social software and web2.0/3.0 applications like social network sites and microblogging services. Herein lie promising challenges and opportunities for CSCL research and development to further evolve in the light of technological advancement and novel forms of interaction. The relevance of web 2.0 dynamics and tools has already been extensively discussed within the CSCL community, particularly in the discourses on collaborative knowledge building (Lu et al. 2010; Scardamalia et al. 2012). Also, the quality of massiveness has been introduced under the term of mass collaboration as one of the future-oriented topics for the design of and the research on collaborative learning environments (Cress et al. 2016; Fischer 2016; Jeong et al. 2017).

This paper contributes to the ongoing discourse. Firstly, we try to conceptualize the socio-technical construct behind the above-mentioned phenomena, that we further refer to as the crowd. We distinguish the crowd from a community and outline it as a distinct entity with specific qualities. In this way, the crowd serves as a productive anchor point to envision innovative learning scenarios and learning environments that are designed to transcend conventional classroom or distant learning settings towards participatory, engaging and fast paced forms of interaction. Secondly, we put forward a design framework for CSCL scenarios that allows to discuss implications of the crowd for the development of a learning environment that supports large groups of learners to establish collaborative knowledge practices. Building on four forms of mediation (Paavola et al. 2012), we introduce five design areas - learner involvement, coordination and regulation, sociality, collaboration and knowledge construction, assessment and feedback - under which we deepen our understanding of the crowd for learning contexts. Finally, we discuss in what way and to what extent the crowd as an organic entity and as an institutionalized and formalized educational setting - as most CSCL scenarios suggest - can be integrated with one another productively. Additionally, forms of crowd inquiry and crowd insight must be specified, especially in comparison to conventional scientific knowledge practices. Thus, this paper provides clarification regarding the current and highly relevant social phenomenon of crowds and systematically opens it up for further theorization and the development of computer supported collaborative learning environments.

## Conceptualizing *the Crowd* for a CSCL context

In the following section, we discuss the crowd from two complementary angles. We ask how the crowd can be understood in its procedural nature - mainly in its formation, its internal coherence and its actions - and how we

can describe it in its fundamental structural and socio-technical properties. To ensure integrability within the current CSCL discourse, we structure our argumentation by building on the state of the art of the mass collaboration approach and by deriving it closely from widespread and acknowledged concepts, namely the Community of Practice (CoP) (Lave & Wenger 1991) and the Community of Interest (CoI) (Fischer 2001; Fischer 2007). As stated above, mass phenomena are quite different in nature. Not only explicit cases of education or, in a broader sense, knowledge creation come to mind but also instances of civic engagement, socio-political participation and activism. While in the field of institutionalized education massiveness remains an ongoing challenge, at the same time we can observe the emergence of informal large-scale accumulations of respective reach, vigor, creative power, productivity and voice. Following, the latter serve as a starting point to conceptualize the crowd and discuss it from a CSCL point of view. Therefore, with the crowd, we are facing a complex social entity, characterized by a highly dynamic nature and intricate internal structure, making use of diverse sets of socio-technical tools and infrastructures and bridging between online landscapes and physical arenas.

From an educational perspective, these current mass phenomena lack a theoretical framework which approaches the crowd as a distinct type of socio-technical entity. Large-scale interactions, in an admittedly expansive understanding, are mainly discussed in connection to the CSCL concept of mass collaboration (Cress et al. 2016; Fischer 2016; Jeong et al. 2017). Despite the marginal conversation about the crowd as a theoretical construct and the use of the term itself, the discussion on how participants come together and engage in large-scale collaborative efforts offers a valuable starting point to establish an understanding of the crowd. Under the term of mass collaboration we are able to distinguish the collaboration in massive scale environments from interactions in small groups or teams. Mass collaboration, in the broadest sense, is described as a large gathering of people in order to work or learn together (Fischer 2016). These social entities are understood as knowledge communities which Jeong et al. (2017) use as an umbrella term to describe a union of people in which knowledge is shared, picked up by others and in some cases newly constructed (Cress et al. 2015). For such knowledge communities, Jeong et al. (2017) list three distinct characteristics in terms of their interaction. Firstly, when such knowledge communities come together at arguably large scale, this form of interaction results in the collaboration among participants from various backgrounds and different geographical locations (Jeong et al. 2017; Fischer 2016). Secondly, despite the differences in knowledge background and geographical location, the participants of knowledge communities manage to organize their work in order to collaborate on artifacts (Jeong et al. 2017). Lastly, the product of mass collaboration usually entails knowledge creation based on collective work efforts (Jeong et al. 2017). Although this definition includes an underlying notion of a crowd, by taking a closer look, it becomes clear that the crowd and the community are not considered separately. Yet, we have reason to believe that the crowd bears its own distinct features which we discuss in more detail in the following.

Knowledge communities understood in the context of massiveness are defined by bearing a shared “mission” (Jeong et al. 2017) in order to achieve a productive way of working together. We argue that when considering the crowd, this shared “mission” has to be viewed from two complementary perspectives. A shared concern on the one hand is what incites masses of participants to form and maintain a crowd while joint efforts on the other hand are their unified actions. To illustrate this distinction: For the social justice crowd phenomenon #blacklivesmatter the unified rage about police brutality against citizens of african-american decent can be regarded as a shared concern and the various activities within the movement (for example using the hashtag across the whole social media landscape or participating in rallies and demonstrations) represent the joint efforts of the crowd. Shared concerns in our understanding of crowds are able to initially and virally vitalize exceptionally large amounts of people and keep their unified momentum going. Whether or not crowds eventually dissolve without a shared concern or with a concern losing its force of attraction over time - just like CoIs dissolve, after a project has ended (Fischer 2007). Another possible scenario could be the gradual stabilization or institutionalization of certain domains of a crowd in more sustainable communities. Also, communities may not only be seen as a remnant of crowd activity but as a relevant factor when it comes to virally spreading a concern or collectively joining efforts. In this reasoning we assume that crowd and community aspects interplay in a certain way for productive large-scale collaboration, the complex synergy of those socio-technical entities has yet to be empirically explored in more detail.

To further distinguish the crowd from communities, we consult the conceptions CoP and CoI. For one thing, a CoP is characterized by a shared social practice as a constituting element. It implies specific properties such as long-lasting, ongoing participation and holding together the members through the attribution of certain roles (e.g., newcomer/oldtimer). The internal structure of a CoP has an underlying notion of being potentially open to anyone while practically offering restricted access for people that can constitute themselves as a member by enacting and upholding the communities’ shared practices or by becoming part of it by progressively adapting (Lave & Wenger 1991). This notion of community is typically used to describe social unions of people from similar (knowledge) backgrounds which are defined by a form of biased communication as the members only



communicate amongst each other, building up on a common background (Fischer et al. 2007, p. 14). Since the crowd as pointed out is mainly held together by its shared concern and joint efforts, the diversity among its participating individuals may reach far beyond that of a CoP. Although CoIs (primarily described as focusing on a specific mutual interest) include the notion of diverse members as well (Fischer 2001), it is important to point out that CoIs fail to pick up the argument of viral temporality and the inherent momentum that we would ascribe crowds. CoIs per definition present a “community of representatives of communities” (Fischer et al. 2007, p. 13) for example in interdisciplinary project teams. In contrast to these conceptions of community, a crowd does not exhibit the procedural characteristics of moving from peripheral to full and core membership (as conceptualized by Lave & Wenger 1991), for example engaging in an ongoing process of enculturation by upholding a shared practice. Rather, in a crowd, various perspectives on shared issues are brought together and expressed side by side, not in a way of successively becoming a part of something by adapting to a set of enacted practices but in a way of being momentarily involved. Certainly shared practices play an important part in connecting the various members of a crowd and securing their mutual understandings to some extent. Nevertheless we would argue that there is a declining significance of inherent processes of establishing convergence for the social entities at hand. Within the concept of CoPs there is a strong focus on reenacting a given internal structure. This is also discussed for CoIs with the difference that a diversity in participants calls for processes of negotiating and establishing common ground to a greater extent (Fischer 2007). For massive-scale interactions however it can be put into question if common ground and a widely accepted consensus provide an underlying mechanism of agreement at all or if antagonisms and notions of dissent can coexist in a unified crowd without endangering the overall efforts.

To this point we have established the crowd by building on a community perspective. To add to this we have to further take into account the crowd not only as a social entity but as a socio-technical one. Crowds are inextricably connected with modern informational and communicational technology. Especially Web 2.0 and 3.0 applications play a significant role when it comes to the aspect of virality, various forms of contributions and the modes of communication within a crowd. In this, crowds are not bound to one service alone but act across a diverse range of social media platforms in particular. Additionally, crowd efforts are not limited to virtual environments but also cover joint activities within offline spaces.

Given the fact that we establish a conceptualization of the crowd against a pedagogical background, naturally the aspect of learning must not be neglected. Therefore, we concern ourselves with the question what an approach to learning in crowd settings could look like. One specific model, that we assume to be adaptable in a crowd context, is the rhizomatic learning approach (Cormier & Stewart 2010). We see a great chance in further refining the concept of rhizomatic learning - which in its core means the distribution of knowledge in digital environments without hierarchical forms of dispersion - in close correlation with our notion of the crowd as a specific case of application.

Concluding, this attempt to provide a first heuristic discussion of the crowd, we can summarize the following procedural and structural qualities: The crowd in our understanding is a mass phenomenon that virally emerges around a concern that touches all participants in their own way and initiates them to contribute to joint actions. The crowd's members are diverse. They face the shared concern from potentially very different perspectives while simultaneously being connected and activated by their smallest intersection, their mutual affectedness. Therefore, the crowd's temporality is solely bound to its own momentum. The crowd is potentially open for everyone to join. Participation involves solely an act of contribution and no process of growing into a shared practice. In this sense, the content of a contribution may differ broadly in regard to the individuals behind them but they are not endless in their variation. Rather, forms of meaningful participation are connected to overstretching practices that are reproduced or emerge within the crowd's socio-material manifestation (e.g., posting on a social network platform), social events (e.g., taking part in a demonstration) and the crowd's operational area (e.g., writing an article for an online encyclopedia). In this, the crowd proposes a promising new take on education in a digital society.

## **Collaboration on a massive scale - a design framework**

Thus far we have proposed a first conceptualization of the crowd that ties in the current CSCL discourse. Now, the next question is how to approach this insight from a developmental perspective to envision innovative CSCL scenarios on a massive scale. More precisely, the question is to which extent learning activities in the crowd raise design issues that resemble or go beyond those, pertinent to other CSCL scenarios and environments.

To identify crucial themes of interest for computer supported collaboration and learning in general, we conducted an exploratory literature review into notable research discourses that bring together notions of collaboration and knowledge creation, namely CSCL research and research on open education. To pinpoint the consequences of massiveness, we then enriched our findings with insights from additional domains that address scenarios of large-scale participation and creation, like citizen science and open source software development.

Within these domains we focused on (1) collaborative scenarios in general while keeping in mind the transferability to, as well as consequences for the crowd context and (2) particular scenarios that involve arguably massive participation.

Table 1: Design Framework for CSCL Scenarios

| <b>Formational Mediation:</b><br>Learner Involvement                      | <b>Pragmatic Mediation:</b><br>Coordination & Regulation   | <b>Social Mediation:</b><br>Sociality  | <b>Epistemic Mediation:</b><br>Collaboration & Knowledge Construction   | <b>Reflective Mediation:</b><br>Assessment & Feedback  |
|---|--|--|---|--|
| Recruiting<br>Participation<br>Engagement<br>Motivation<br>Sustainability | Dynamic<br>Static<br>Intrinsic<br>Extrinsic<br>Process awareness<br>Decision making<br>Didactics<br>Technology<br>Negotiation<br>Established practice<br>Hierarchy | Communality<br>Social awareness<br>Communication<br>Authorship<br>Accountability<br>Responsibility | Common ground<br>Resources<br>Knowledge objects<br>Knowledge artifacts<br>Shared insights<br>Currents of inquiry<br>Epistemic frames<br>Subjects<br>Social permeability | Many-to-many<br>Peer-to-peer<br>Instructor-based<br>Creation of value<br>Institutionalization<br>Formalization |

To be able to map and compare critical elements of crowd-based learning scenarios in relation to other CSCL settings, we build on the multidimensional model of mediation, suggested by Béguin & Rabardel (2000) and adapted to the field of CSCL by Paavola et al. (2012). According to Paavola et al. (2012), CSCL scenarios can be described and analyzed along four dimension or types of mediation. These types of mediation include (a) epistemic mediation, the procedural means through which knowledge is created collaboratively, (b) pragmatic mediation, the procedural means through which the collective efforts are coordinated and regulated, (c) social (or collaborative) mediation, the procedural means through which social relations and networks are (re-produced), as well as (d) reflective mediation, the procedural means through which the collective efforts are assessed, evaluated and advanced. The literature review however pointed to a set of design issues, which goes beyond the types of mediation suggested by Paavola et al. (2012). These issues relate to what Jeong et al. (2017) have discussed as different “forms of joint interaction that occur in large-scale community settings” (p. 134). We take up the relevance of categorizing such phenomena to describe how people work together and interact with each other by abstracting it to a more nuanced idea of learner involvement which is concerned with those processes and means by which actors are recruited and participation is motivated and ensured. We subsume these issues under the notion of (e) formational mediation, the procedural means through which involvement and participation are ensured. Table 1 provides an overview of the design issues we have identified with regard to these five types of mediation. The table can be understood as generic framework for the description, analysis, design and evaluation of computer supported collaborative learning scenarios and environments. In the following, we will use this framework as a vantage point to further discuss the particularities of crowd-based learning scenarios.

### Formational mediation – Learner involvement

The first question for the setup of collaborative learning is that of how to address and involve participants and who those participants are. The recruiting of learners, so to say. Thinking along the lines of a crowd setting that transcends a given and formalized union of a classroom, recruiting implies that a critical mass of individuals has to be attracted, for crowd dynamics to develop at all. The actual process of rallying a crowd can only be understood in regard to the party responsible for the recruitment - for example a research team in a citizen science context (Chu et al. 2012), a MOOC provider (Haywood 2016, p. 75) or the crowd itself in its own momentum. Following this argument, the internal structure of a crowd comes into view, since different activities call for different levels of diverse or shared perspectives or practices among the involved individuals (Bonney et al. 2016) - regardless of the externally given or internally established origin of mutual objectives. Therefore, the participants and their individual qualities can be of high importance as well, because for being successfully involved in the joint effort they might need to have certain knowledge and competences at their disposal (Kobori et al. 2016) or their contributions need to achieve a specific standard to be considered of value or compatible at all (Ghosh & McAfee 2011). Adding to the consideration of recruitment, a distinction between participation and engagement seems reasonable. Within the frame of participation we can ask in what ways participants are able to take part in a meaningful way, which forms of contribution are compatible and if members face some kind of participational

threshold. Engagement on the other hand addresses the ongoing and active involvement of participants and is therefore a vital topic within the open education and open science discourse (Walji et al. 2016; Sprinks et al. 2017). Considering the social entity of a crowd we must challenge the idea that ongoing and active involvement constitutes a desirable participant. On the contrary, we should be mindful of punctual - and maybe minimal - acts of contribution and dissemination in consideration of the crowd's shared efforts and concern.

Following this contemplation, the view on motivation changes as well. Participant motivation is a leading topic for all kinds of educational and participatory scenarios (Rogat et al. 2013; Eveleigh et al. 2014) while the explicit implications for crowd-based learning have yet to be discussed. In the light of heterogeneous participants, diverse forms of participation and various modes of engagement as well as shared concerns and efforts as a constituting element of a crowd, motivation has to be conceptualized between these contrasting poles of enormous variability of being part of a crowd and presumably more common reasons to be involved at all.

Finally, we have to question the live-cycle of a crowd, for example its sustainability. With the constitutive factor of a shared concern and a joint effort, a crowd may dissolve when those inherent impulses are no longer given or may partially evolve in more stable forms of communality like a community or an organization. Besides this, processes of institutionalization of crowd phenomena, the compatibility of temporal arbitrariness of crowds and often temporally ridged educational structures have to be considered when imagining learning in a crowd-based environment.

### Pragmatic mediation – Coordination and regulation

The various ways in which collaborative processes are coordinated and regulated are an ongoing issue within CSCL discourse (Ludvigsen et al. 2018). Assuming an inseparable relation between didactic prompts and interventions and technical affordances, we have to ask how participants are able to interact with the environment and with each other in a meaningful way. Especially when designing for a crowd the question arises how didactics and technical guidelines can support the joint efforts without restricting or preventing the crowd's constitutive momentum. In short, how can an emergent property like virality be facilitated at all? It is also relevant to consider the coordinating and regulating aspects of a learning scenario between the poles of static and dynamic structures and intrinsic and extrinsic determinations. Is a structure for example pre-established or does the environment enable the participants themselves or an external instructor to question, adapt and change their ground rules of collaboration? With the quantity and the diversity of crowd participants in mind, we further have to ask if equal negotiation and unanimous decision making is possible or even necessary for the crowd to be able to act and for people to join. Rather it seems essential to establish some kind of process awareness to help participants to find ways for at least minimal or perhaps more elaborate forms of contribution. Process awareness here serves as an ideal concept, because within the crowd's complexity and its emergent evolution it is an impossible task to concurrently describe it entirely. Finding the fine line to foster a crowd's dynamic without crippling it with a tight pedagogical frame and ensuring access for learners despite its complexity seems to be one of the main challenges when designing crowd scenarios for educational context.

### Social mediation – Sociality

We have outlined the crowd as a socio-technical entity that is on the one hand diverse, complex and to an extent unbound and on the other hand temporally constituted under a joint concern and with a mutual cause. By accommodating such qualities, it can be clearly distinguished from other forms of communality that are addressed in CSCL scenarios as we argued in the previous paragraph.

Another concept that relates to sociality within the fields of computer supported collaboration is that of social awareness (Dourish & Bellotti 1992). It addresses possible ways for users in a virtual environment to perceive themselves as part of a social unit and consequently take note of each other. In this regard we raise questions at three levels. How is an individual able to become aware of all the other participants in the crowd entity, given the fact that so many people are involved in different ways at different levels (Jeong et al. 2017)? How does a notion of being part of a mass-movement emerge and what shapes this experience? And lastly, how can social awareness as a concept address the crowd as an entity that may not be attached to one specific environment but stretches over multiple virtual spaces and spreads into the offline world as well?

When members of a crowd communicate with each other, it can reasonably be concluded that communication exceeds dialogical forms. Such exceeding communication modes - which we assume to be highly relevant for crowd settings - can be observed in social networking or microblogging applications (e.g., in follower-follower relations, under shared hashtags or through @mention functionalities). With those mass-environments in mind, communication modes like one-to-many broadcasting (Page 2012) or pushing of pre-existent content without adding to it (Kaplan & Haenlein 2011) have to be considered in their consequence for the development of means of communication within learning scenarios for the crowd.

Another important factor that has to be thought anew in crowd settings is that of authorship which also implies the question of accountability and responsibility. We assume that in part, participants maintain visibility as authors as such. Additional scenarios though are anonymous forms of participation as well as rather content-less acts of meaningful participation (e.g., liking or reblogging/retweeting) that are immanent in crowds. In this case the question arises who can be held responsible for contents, statements and actions. If we consider a joint effort of a crowd as manifestation of its work, is it even possible to ascribe authorship, responsibility or accountability to a single person? Or must a crowd as entity be held accountable or responsible somehow?

### **Reflective mediation – Assessment and feedback**

Assessment and feedback structures are among the most recurring and discussed issues in the CSCL community, not just as final evaluation method, but especially as an advancing act within the process of collaborative learning. In this regard the concept of peer-to-peer interaction is of particular importance and investigated in its own qualities (Reinholz 2016) as well as in its differentiation from instructor-based feedback and assessment mechanics (Harney et al. 2017). Additionally, many-to-many assessment scenarios have been explored within the field of open education and MOOC development, with the expectation of bringing together different levels of expertise, experience and prospects to evaluate contributions (Clougherty & Popova 2013). For crowd-based collaboration we have to ask how hierarchy-driven, peer-driven and in particular crowd-driven feedback and assessment dynamics take shape and in which way they can be fostered or are emerging organically during massive-scale interactions. In the open, complex and diverse socio-technical structure of the crowd, the ways in which the ascription of value unfolds in regard to the joint concern and efforts can therefore only be anticipated as an issue of future investigation. We assume that in addition to evaluating and reflecting functions of assessment and feedback, we will observe aspects of admission, such as generally including contributions as meaningful or disregarding them as irrelevant. In comparison to institutionalized educational settings, we assume that there are no explicit and formalized forms of valuation (e.g., examinations, grades or standards of education) existent in the crowd. Nevertheless, we ask ourselves if processes of formalization - beyond the shared practices of participation - play a role in crowd environments and how they can be explored further.

### **Epistemic mediation – Collaboration and knowledge construction**

Lastly, we have to consider the crowd from an epistemic standpoint and discuss the circumstances of collaborative knowledge practices. To understand such processes, we have to determine the crowd's resources which ultimately create the foundation for knowledge construction. These resources include a certain diversity in perspectives, different approaches to problem solving and by sheer mass of participants, a major workforce.

Within a heterogenous crowd we can assume participants from various epistemic backgrounds to bring in their disciplinary practices and standpoints or, as Shaffer (2006) suggests, their epistemic frames. For a crowd setting it has to be revised, if an epistemic diversity only presents itself as a resource for multifaceted knowledge creation or if it may interfere as a participatory obstacle or even a gatekeeper. Diversity may call for the negotiation of common ground. But, as we have already discussed that common ground in a crowd context is sufficiently ensured by the shared concern and that beyond that, integrability of contributions only arises by the participatory act of contributing itself. Thus, different perspectives and approaches can coexist under a shared concern and are realized through joint efforts without necessarily coming into conflict with one another.

Framing crowd activities as an epistemic endeavor the question arises what can be considered as the outcome of a collaborative knowledge work. What shapes do crowd insights might take and are they for example of a more collective or a more individual nature. Clarification is needed on how such insights can become visible at all. How can a multitude of contributions together form a coherent current of inquiry? To further elaborate on the shared concern from an epistemic perspective, we see great potential in adapting Knorr Cetina's (2001) concept of Knowledge Objects and partial artefacts. Especially the notion of partiality we see as a starting point to examine the role of scattered contributions in relation to the collective epistemic process of a crowd.

Finally, recalling the observation that crowd phenomena are socially motivated, the matter of permeability comes into focus. We have to ask in what way integrability beyond the educational context of crowd insights can be established not only for scientific societies but for civic discourses. This can be understood in two different ways: For one, the crowd can be enriched and motivated by sources beyond the specific educational context, but also, its insights can be regarded in their societal accessibility.

### **Implications and discussion**

With this paper we have put forward two contributions to the consideration of massiveness within CSCL contexts. Building on the discourse around mass collaboration and the concepts of CoPs and CoIs, we have conceptualized the crowd as a first attempt to establish a productive connection between current mass phenomena and educational

design. Within a framework of five areas of mediation we then took this conception of the crowd as a vantage point to raise issues of collaborative mass scenarios in a more detailed way. Drawing conclusions from this approach, we see the crowd as an opportunity to evolve computer supported collaborative learning alongside current technical and societal processes of transformation. Nevertheless, or possibly for exactly this reason, the crowd presents itself as a challenging design target.

We have outlined the crowd as a highly complex entity with an emergent nature and an inner momentum that is primarily fueled by the mutual affectedness of its participating individuals. The open question remains how an educational design can address all those issues in a way that fosters a crowd's formation and productive interactions within the scope of pedagogical expectations and constraints without inadvertently suppressing the dynamics. How can we achieve to cultivate an organic entity inside institutionalized education? Or do we rather have to envision a new form, a guided crowd, as a formalized variety alongside informal crowd phenomena?

In the light of the crowd, the aspects of learning, of collaboration and of inquiry require a reconsideration as well. We can ask where learning processes arise while taking part in crowd activities and which understanding of learning seems suitable to describe and facilitate those instances (e.g., rhizomatic learning). Further, we may have to widen our understanding about subjects of learning all together and consider not only the individual learner but the crowd as a collective learning entity. It is also worthy of discussion whether or not present conceptions of collaboration are suitable to describe joint efforts where interpersonal activity is not a constituting circumstance, or if crowd phenomena can lead to an extended theorization of collaboration all together. Another fundamental challenge lies within the CSCL objective to foster forms of inquiry derived from scientific practice. An idea of crowd-based inquiry may help to broaden the perspective of meaningful collective knowledge creation, especially in close relation to societal and cultural issues.

Finally, to entirely grasp the crowd, we have to acknowledge it not only as a socio-technical entity but in its performativity. The crowd as such only exists as long as individuals engage and contribute in a certain way. Only from this joint action the crowd emerges and is produced and reproduced in its particular features and only by contribution, participants can perceive themselves as being a part of it. Thus, understanding the crowd as a performative act is marked by an inherent contingency that proposes highly relevant implications not only for the design of educational crowd scenarios but for their scientific exploration as well.

## References

- Béguin, P. & Rabardel, P. (2000) Designing for instrument-mediated activity. *Scandinavian Journal of Information Systems*, 12, 173-190.
- Bonney, R., Phillips, T.B., Ballard, H.L., & Enck, J.W. (2016b). Can citizen science enhance public understanding of science? *Public Understanding of Science*, 25(1), pp. 2-16.
- Chu, M., Leonard, P. & Stevenson, F. (2012). Growing the base for citizen science. In: Dickinson, J.L. & Bonney, R. (Eds.), *Citizen science: public participation in environmental research*. Cornell University Press, Ithaca, pp. 69-81.
- Clougherty, R. J. & Popova, V. (2013). Crowdsourcing to Assess MOOCs: A Position Paper. *MOOCs FORUM*, 1(P), pp. 10–13.
- Cormier, D., & Stewart, B. (2010). Life in the open: 21st Century learning & teaching. *Atlantic Universities' Teaching Showcase*, 24, pp. 24-31.
- Cress, U., Stahl, G., Ludvigsen, S. & Law, N. (2015). The core features of CSCL: Social situation, collaborative knowledge processes and their design. *International Journal of Computer-Supported Collaborative Learning*, 10(2), pp. 109-116.
- Cress, U., Jeong, H. & Moskaliuk, J. (2016). Mass Collaboration as an Emerging Paradigm for Education? Theories, Cases, and Research Methods. In U. Cress, J. Moskaliuk & H. Jeong (Eds.), *Mass Collaboration and Education* (pp. 3-27). Cham: Springer International Publishing.
- Dourish, P.; Bellotti, V. (1992): Awareness and coordination in shared workspaces. In: M. Mantel und R. Baecker (Hg.): *CSCW '92 Proceedings of the 1992 ACM conference on Computer-supported cooperative work*. New York, pp. 107-114.
- Eveleigh, A., Jennett, C., Blandford, A., Brohan, P. & Cox, A.L. (2014). Designing for Dabblers and Detering Drop-outs in Citizen Science. In: *Proceedings of the 32<sup>nd</sup> Annual ACM Conference on Human Factors in Computing Systems, CHI '14*. ACM, New York, NY, USA, pp. 2985-2994.
- Fischer, G. (2001). Communities of Interest: Learning through the interaction of multiple knowledge systems. In: S. Bjornestad, R. Moe, A. Morch, & A. Opdahl (Eds.), *Proceedings of the 24th IRIS Conference*, pp. 1-14. Bergen: Department of Information Science.

- Fischer, G., Rohde, M., & Wulf, V. (2007). Community-based learning: The core competency of residential, research-based universities. *International Journal of Computer-Supported Collaborative Learning*, 2(1), pp. 9-40.
- Fischer, G. (2016). Exploring, understanding, and designing innovative socio-technical environments for fostering and supporting mass collaboration. In: Cress, U., Moskaliuk, J. & Jeong, H. (Eds.), *Mass collaboration and education*, pp. 43-63. Cham: Springer International.
- Ghosh, A. & McAfee, P. (2011). Incentivizing high-quality user-generated content. In: *Proceedings of the 20th International conference on World Wide Web*. ACM, pp. 137-146.
- Harney, O. M., Hogan, M. J., & Quinn, S. (2017). Investigating the effects of peer to peer prompts on collaborative argumentation, consensus and perceived efficacy in collaborative learning. *International Journal of Computer-Supported Collaborative Learning*, 12(3), pp. 307-336.
- Haywood, J. (2016). Learning from MOOCs: lessons for the future. *Wenner Gren Symposia*, 88(6), pp. 69-79, Portland Press.
- Jeong, H., Cress, U., Moskaliuk, J. & Kimmerle, J. (2017). Joint interactions in large online knowledge communities: The A3C framework. *International Journal of Computer-Supported Collaborative Learning*, 12(2), pp. 133-151.
- Kaplan, A. M.; Haenlein, M. (2011): The early bird catches the news: Nine things you should know about micro-blogging. In: *Business Horizons* 54 (2), S. 105–113.
- Knorr Cetina, K. (2001): Objectual practice. In: T. R. Schatzki, K. Knorr Cetina und E. & von Savigny (Hg.): *The Practice Turn in Contemporary Theory*. New York: Routledge, S. 184–197.
- Kobori, H., Dickinson, J.L., Washitani, I., Sakurai, R., Amano, T., Komatsu, N., Kitamura, W., Takagawa, S., Koyama, K., Ogawara, T. & Miller Rushing, A.J. (2016). Citizen science: a new approach to advance ecology, education, and conservation. *Ecological Research*, 31(1), pp. 1-19.
- Lave, J. & Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. New York: Cambridge University Press.
- Lu, J.; Lai, M.; Law, N. (2010): Knowledge Building in Society 2.0: Challenges and Opportunities. In: Khine M. S. und Saleh I. M. (Hg.): *New Science of Learning. Cognition, Computers and Collaboration in Education*. New York: Springer, pp. 553-568.
- Paavola S., Engeström R., Hakkarainen K. (2012). The Triological Approach as a New form of Mediation. In: Moen A., Mørch A.I., Paavola S. (Eds.) *Collaborative Knowledge Creation. Technology Enhanced Learning*, vol 7, pp. 1-14. SensePublishers, Rotterdam.
- Page, R. (2012): The linguistics of self-branding and micro-celebrity in Twitter: The role of hashtags. In: *Discourse & Communication* 6 (2), S. 181–201.
- Reinholz, D. (2016). The assessment cycle: a model for learning through peer assessment. *Assessment & Evaluation in Higher Education*, 41(2), pp. 301-315.
- Rogat, T. K., Linnenbrink-Garcia, L., & DiDonato, N. (2013). Motivation in collaborative groups. *International handbook of collaborative learning*, 250-267.
- Ross, J., Sinclair, C., Knox, J., Bayne, S. & Macleod, H. (2014). Teacher experiences and academic identity: The missing components of MOOC pedagogy [MOOC Special issue]. *MERLOT Journal of Online Learning and Teaching*, 10(1), pp. 57-69.
- Scardamalia, M.; Bransford, J.; Kozma, B.; Quellmalz, E. (2012): New Assessments and Environments for Knowledge Building. In: P. Griffin, B. McGaw und E. Care (Hg.): *Assessment and Teaching of 21st Century Skills*. New York: Springer, pp. 231-300.
- Shaffer, D.W. (2006). Epistemic frames for epistemic games. *Computers & Education*, 46(3), 223–234.
- Sprinks, J., Wardlaw, J., Houghton, R., Bamford, S., & Morley, J. (2017). Task Workflow Design and its impact on performance and volunteers' subjective preference in Virtual Citizen Science. *International Journal of Human-Computer Studies*, 104, pp. 50-63.
- Walji, S., Deacon, A., Small, J., & Czerniewicz, L. (2016). Learning through engagement: MOOCs as an emergent form of provision. *Distance Education*, 37(2), pp. 208-223.

## Acknowledgments

The research leading to these results has received funding from the Federal Ministry of Education and Research (BMBF) under grant agreement n° 16DHB2119 and was carried out as part of the project: »SCoRe – Student Crowd Research«. The content of this article does not necessarily reflect the opinion of the Federal Ministry of Education and Research.

# Adaptive Support for Collaboration on Tabletop Computers

Abigail C. Evans, Khoury College of Computer Sciences, Northeastern University, ab.evans@northeastern.edu  
Katie Davis, Information School, University of Washington, kdavis78@uw.edu  
Jacob O. Wobbrock, Information School, University of Washington, wobbrock@uw.edu

**Abstract:** We present the design, implementation, and evaluation of a system providing adaptive support for collaborative learning at a tabletop computer. Adapting to support collaboration involves tackling several sub-problems: detecting when a group is struggling, determining when the system should provide support, what sort of support it should provide, and for how long. A classroom evaluation of our system showed it was effective at detecting collaboration breakdowns and provided preliminary evidence that just-in-time adaptive support for collaboration might help to deter disruptive behavior in small group settings.

## Introduction

Computer-supported collaborative learning (CSCL) environments are becoming increasingly sophisticated, enabling new ways for students to work together. However, prior research has demonstrated that students do not always know how to collaborate effectively, which can inhibit the success of small group learning, e.g., (Dillenbourg & Jermann, 2007; Järvelä & Hadwin, 2013; Rogat & Linnenbrink-Garcia, 2011). The findings of this prior work suggest that collaboration itself is a skill that needs to be developed in the classroom.

In this work, we focus on face-to-face, small group collaborative learning at tabletop computers because their large, shared touch screen makes them well-suited for collaboration. However, the mere presence of a large touch screen cannot make up for a lack of collaboration skills—students who do not know how to collaborate effectively in traditional face-to-face settings will continue to struggle at a tabletop computer. Our long-term goal has been to enable tabletop computers to make up for poor collaboration skills. Existing work has addressed the first step in realizing this goal, namely detecting collaboration breakdowns (Evans, Wobbrock, & Davis, 2016; Evans & Wobbrock, 2014). In this paper, we present our work toward the next step—designing ways for the tabletop interface to adapt when breakdowns are detected in order to encourage more effective collaboration. We describe the design and implementation of our system, which was guided by learning theory and educational research, and a classroom evaluation. Our analysis focuses on how the configuration of the adaptations appeared to influence group dynamics over time and through repeated exposure for one group of students particularly prone to collaboration problems. Our key findings are that the adaptations deterred disruptive behavior and halved the length of periods of low-quality collaboration for this group.

## Measuring collaborative learning

To determine what qualifies as *effective* collaborative learning, we use the concept of social regulation, which refers to “the social processes groups use to regulate their joint work on a task” (Rogat & Linnenbrink-Garcia, 2011). Social regulation is an extension of the concept of self-regulation in individual learning to groups of learners in a collaborative setting (Järvelä & Hadwin, 2013; Panadero & Järvelä, 2015; Rogat & Linnenbrink-Garcia, 2011; Volet, Vauras, & Salonen, 2009). Social regulation typically refers to metacognitive processes independent of domain knowledge, although this relationship needs further research.

Rogat and Linnenbrink-Garcia (2011) identify three dimensions of social regulation: *planning* the group’s approach to a task, *monitoring* of understanding and progress, and *behavioral engagement*—efforts to get group members to engage with the task. A group’s use of each dimension can vary in quality, with high-quality social regulation processes leading to socially-shared regulation, in which all group members maintain joint attention on the learning activity, regularly checking in on the goals, plans, and progress. Low-quality social regulation processes, such as when groups fail to come up with an appropriate plan, can lead to poor learning outcomes. We chose to follow Rogat & Linnenbrink-Garcia’s framework when coding for social regulation. Their approach was selected above others because of their detailed descriptions of what high- and low-quality social regulation processes look like in practice; thus, it could easily be applied to our own work.

## Adaptive support for collaborative learning

There has been extensive prior work on adaptive support for collaboration in text-based CSCL environments. For example, guiding systems coach students through a collaboration using adaptive feedback, e.g., (Hadwin, Oshige, Gress, & Winne, 2010; Kumar & Rosé, 2011; Wang, Rosé, & Chang, 2011), and group awareness tools provide feedback to students, often in the form of visualizations, on how they are collaborating, e.g., (Järvelä et

al., 2014; Malmberg, Järvelä, Järvenoja, & Panadero, 2015; Trausan-Matu, Dascalu, & Rebedea, 2014).

Many of the group awareness tools described in the literature make use of the closed nature of text-based collaborative environments, in which all group interactions can be captured and mined for patterns associated with the quality of collaboration. In face-to-face group work at a tabletop, however, the computer can only access direct human-tabletop interaction—the verbal and gestural interactions that learners have with each other cannot be captured without the use of external sensors. Martínez-Maldonado et al. used external sensors to capture learners' verbal and physical interactions during tabletop collaboration in order to visualize each group member's level of participation (Martínez-maldonado, Kay, Yacef, Edbauer, & Dimitriadis, 2013; Martínez-Maldonado, Yacef, Kay, Al-Qaraghuli, & Kharrufa, 2011; Martínez, Collins, Kay, & Yacef, 2011). Although these visualizations were aimed at teachers, they could also be presented to students as group awareness tools. At the time of this writing, the work by Martínez-Maldonado et al. remains the only other research providing real-time adaptive support for collaborative learning at tabletop computers.

It is important to consider how adaptive support should be presented to students and situated in the larger context of a learning activity or course. Wise (2014) notes it is not enough to simply present students with interventions—if students are to benefit from an intervention, they need to understand its intent and how to engage with it. An intervention should be an “integral part of course activity tied to goals and expectations” (Wise, 2014, p. 206). Interventions should also encourage students to have agency in their own learning.

## System design

Providing real-time adaptive support for collaboration involves two major components: (1) *detection* of behavior that will trigger adaptations, and (2) the software *adaptations* themselves. A third consideration is when and how adaptations are triggered, which we refer to as the *configuration*.

### Detection

We used Evans et al.'s approach to detecting the quality of collaboration at a tabletop computer using a group's touch-interaction patterns (Evans et al., 2016). Evans et al. identified two touch patterns that serve as reliable indicators of a group's collaboration quality: *unrelated touches* (UT), and *overlapping unrelated sequences* (OUS). A *sequence* is a series of touches carried out by an individual that represents a complete action. UT measures the proportion of touches in a sequence that involve objects that are unrelated in the context of a learning activity, with a greater proportion of touches to unrelated objects corresponding to lower quality collaboration—specifically, off task behavior or failure to settle on a suitable plan to complete the activity. OUS is a measure of whether a single student or multiple students are interacting with the tabletop, and when multiple students are interacting, whether they are working with objects that are related to each other in the context of the activity. Evans et al. found that, during periods of high-quality collaboration, students either took turns to interact with the computer or multiple students worked together with related objects. Multiple students interacting with unrelated objects was a sign that students were either working independently or off-task.

Evans et al.'s detection approach was built in to our learning software following the implementation details described therein (Evans et al., 2016)—two-minute intervals of touch data are checked against the two touch patterns, UT and OUS, each of which returns a label estimating the quality of collaboration occurring in that two-minute period as either high-, medium-, or low-quality. It is important to note that the touch patterns need to be able to distinguish the touches of individual students from those of other group members, which most tabletop computers are unable to do. We also therefore implemented Evans et al.'s *Group Touch* method of distinguishing among tabletop users to meet this requirement (Evans, Davis, Fogarty, & Wobbrock, 2017).

### Adaptations

We aimed to create adaptations that would be application-independent in principle. Therefore, we developed the adaptations without a specific application in mind. Initial brainstorming resulted in 11 adaptation ideas, of which four were implemented and pilot tested. See Figure 1, below, for screen shots of these adaptations:

The *group awareness icon* draws on prior work on group awareness tools e.g., (Järvelä et al., 2014; Malmberg, Järvelä, Järvenoja, & Panadero, 2015; Trausan-Matu, Dascalu, & Rebedea, 2014). An icon remains visible on screen for the duration of the activity and changes color according to the detected collaboration quality, with green representing high-quality collaboration, amber representing medium-quality collaboration, and red representing low-quality collaboration.

*Control lockout* blocks or disables select controls, such as buttons or menu items, to reduce the number of active application elements. The rationale for this adaptation comes from prior work that found that when group members jump around between many different controls, they are typically either off task or struggling to identify or stick to a task plan (Evans et al., 2016). The aim is to encourage sustained whole-group focus on a



specific piece of the assigned activity.

*Voting* prevents actions that affect the global state of the application, such as deleting work or marking an activity complete and moving on, from being carried out unless the majority of group members vote to proceed. The goal is to encourage discussion and prevent individual students from dominating the group.

*Escalate to authority* sends an alert to the teacher’s mobile device requesting human intervention.



**Figure 1.** The adaptations in the study applications. Top left: the group awareness icon, positioned in the corners of the screen, showing amber (medium-quality collaboration). Top right: the prompt, added after the pilot. Bottom left: control lockout has disabled a control that allows students to switch the task. Bottom right: voting blocks actions from being carried out until a majority of the group members vote for it.

### Configuration

The adaptations aim to scaffold effective collaborative interactions by making it difficult for students to interact with the screen in ways deemed to be ineffective. However, no single adaptation covers the full range of desirable collaborative behaviors, all of which need to be adopted by the group if they are to work together effectively. Therefore, we decided that the adaptations should be triggered in a sequence, where each new adaptation is added to those already in use but is only triggered if the collaboration quality does not improve. Layering adaptations can make it easier for students to ground themselves in the activity’s expectations, because together the adaptations give a more complete picture of how to be effective than any isolated adaptation. When the collaboration does improve, all active adaptations are removed at once—an example of “fading,” which, in the scaffolding literature, refers to the process of removing supports that are no longer needed (Pea, 2004).

Removing adaptations when collaboration improves also supports Wise’s (2014) principle of agency. After a group has experienced a sequence of adaptations once, they will know what is coming next if the sequence begins again. Because the students can anticipate the next step in the sequence, they also have the opportunity to preemptively adopt the behaviors that the next adaptation will enforce, removing the need for that adaptation and thereby preventing it from being triggered. For example, if students know that the next adaptation to be triggered if their collaboration does not improve will be *control lockout*, they can aim to improve by interacting only with objects directly relevant to the aspect of the task they are working on.

The sequence begins with the *group awareness icon*, which remains visible on-screen for the whole activity, changing color according to the detected collaboration quality. The second adaptation, *control lockout*, is triggered when the icon turns red, representing poor collaboration quality, and no other adaptations are active. *Control lockout* was chosen as the first restrictive adaptation based on observations from prior work on tablet collaboration that groups often jumped into an activity without taking time to plan or check the instructions (Evans et al., 2016). If collaboration does not improve, the third adaptation, *voting*, is added. *Voting* can slow down progress and become frustrating, so it is only triggered after students have had the opportunity to reflect on and improve their collaboration. Finally, if collaboration still does not improve after a period of voting, the application will *escalate to authority* and call the teacher to intervene. There are many other ways to combine,

order, and layer the selected adaptations in a sequence. This particular sequence was chosen initially as the adaptations progress in order from least to most restrictive or interventionist, giving students time to reflect on and change their collaboration behavior before increasing the level of restrictions imposed on the group.

In addition to deciding which adaptations to build and the order in which they should be triggered, we had to decide exactly when an adaptation should be triggered, how it should be presented and explained to students, how long each adaptation should remain active, and when adaptations should be removed. Each of these choices could impact the outcome of the intervention. The criteria we chose for triggering adaptations were derived from Evans et al.'s (2016) analysis of the reliability of using the touch patterns (UT and OUS) to detect collaboration quality. Adaptations were triggered immediately if both touch patterns labeled a single interval low quality, or if two consecutive intervals were labeled low or medium quality by UT. Adaptations remained active until both touch patterns produced high quality labels for four consecutive minutes.

## Classroom evaluation

The prototype adaptations were evaluated in a classroom setting. Our research questions were: (1) Are the adaptations triggered appropriately? (2) How do students respond when an adaptation is triggered?

## Participants and apparatus

Eleven 10<sup>th</sup> and 11<sup>th</sup> graders (9 male, 2 female) participated in the study. The students were enrolled in a six-week beginner's course on video game development offered as part of a college preparation program serving low income students. The course used in this study was one of the program's elective options. The course focus was educational games, and students were asked to design a game to raise awareness of an environmental issue: snow leopard conservation. In the first part of the course, the students learned about snow leopards and ongoing efforts to protect the species. In the second part of the course, the students learned the basics of video game development using Unity (a 3-D game development platform: see <https://unity3d.com/>). All but one of the students were new to writing code and all were new to video game development and Unity. In the third and final part of the course, the students worked on their game projects.

The students were randomly assigned to three small groups—two groups of four and one group of three. The students stayed in the same groups for all collaborative activities. The groups used three custom-built learning applications across four sessions at a Microsoft Surface Hub, which features a 55" multitouch screen. The applications addressed the learning objectives for the class sessions in which they were to be used and were designed to be used alongside other activities.

*Snow Leopards 101* is an introduction to snow leopards and their ecosystem adapted from an existing non-tabletop curriculum (Facing the Future & Snow Leopard Trust, 2009). In *Help a Scientist*, students take on the role of scientists studying wild snow leopards in Mongolia, gathering and analyzing photographic data. *Game Challenge* was used during the third part of the course. This application featured a partially created game world that students could build upon and adjust to create different outcomes while learning important and complex concepts used in Unity, such as how to work with its built-in physics engine.

The Surface Hub is primarily intended to be used in a vertical orientation but for this study it was placed flat on a table in a small breakout room across a hallway from the main classroom. A wide-angle video camera was mounted on a wall so that students could freely move around the table. The camera was angled toward the screen so that it could capture every touch and the interactions among the group members. Ten group sessions at the tabletop were video recorded for this study and the computer logged every touch.

## Study design

The study was split into three phases: (1) baseline data collection; (2) a pilot test of the initial implementation of the adaptations, which resulted in a revised implementation; and (3) an evaluation of the final implementation. The *Snow Leopards 101* application was used for baseline data collection, *Help a Scientist* was used for the pilot test, and *Game Challenge* was used for the final evaluation taking place two and a half weeks after the pilot. Students used each application for 25 minutes per session. *Game Challenge* was used over two sessions.

The adaptations were piloted with Group 1 only. The final adaptations were available for Group 1 and Group 3 during the final evaluation—if poor quality collaboration were detected, the adaptations would trigger. Group 2 served as a control so the adaptations were not available for them during phase 3, although the collaboration quality labels output by the touch patterns were still recorded. We wanted to look for differences in how groups dealt with collaboration problems with or without the intervention of adaptations. Students were randomly assigned to groups and groups were randomly assigned to conditions in an attempt to reduce differences between groups. However, given that it was unlikely the groups would be truly equal in a study of this scale, the goal of the baseline data collection was to understand the differences between groups.

The first time a group used the tabletop with adaptations enabled, a researcher explained to them that the computer was tracking how they interacted with it in order to help them work together effectively. They were told that the color of the *group awareness icon* updated every minute to give them feedback on how they were collaborating and that if it turned red or stayed amber for several minutes, the computer would block some controls or ask them to vote before carrying out certain actions. They were also told that the adaptations would go away if the icons stayed green for four minutes, and that they could keep the icons green by maintaining a shared focus on the assigned task, discussing the content, and listening to each other's ideas.

## Pilot test

The adaptations were piloted with Group 1, a group of four boys randomly assigned to the pilot, as they used the *Help a Scientist* application. The *group awareness icon* (Figure 1, top left) was displayed in two of the corners of the screen so that it would be visible to all group members without obstructing the work area. *Control lockout* (Figure 1, bottom left) was configured to trigger if the icon turned red when no other adaptations were active. Before activation, a warning popped up on screen stating, "Some buttons and controls will be temporarily disabled to help you to stay focused on the task." The warning remained on screen for 30 seconds, blocking all interaction for the time that it was on screen. *Voting* (Figure 1, bottom right) was also preceded by a 30-second warning message: "You will temporarily be asked to vote in order to carry out certain actions such as changing activity or closing windows." If a group was asked to vote on an action, a message was shown describing what the group members were voting on. Finally, *escalate to authority* would send an alert to the teacher that the group might be struggling. Unlike *control lockout* and *voting*, *escalate to authority* was invisible to the students.

To evaluate the piloted adaptations, the first author reviewed the video of the pilot session and the computer's log files. Given that the purpose of the pilot was to get feedback on a number of design choices quickly enough to make changes in time for a formal evaluation two and a half weeks later, it was not possible to do a full in-depth analysis of the video at this point. Instead, observations were made of what the group was doing in the run up to an adaptation being triggered and how group members responded. This informal analysis was confirmed by a formal analysis of the pilot once the study was complete.

The sequence of adaptations triggered twice during the pilot: at 8 minutes into the session and again at 23 minutes. In both cases, the sequence progressed from *control lockout* to *voting* and was canceled before *escalate to authority*. This means that low-quality collaboration was still detected after at least one interval of *control lockout*, causing *voting* to be triggered. The group was able to sustain high-quality collaboration for at least three intervals while *voting* was active, so the adaptations were removed without messaging the teacher.

After reviewing the verbal and physical interactions among the students, we determined that the adaptations triggered appropriately and that they did appear to get the group to collaborate more effectively. However, the positive behavior change that occurred seemed to be a result of coercion—the adaptations proved so annoying that they forced the students to improve their working style without encouraging reflection. The main lesson learned was that it was too difficult to get the adaptations removed. Students did initially improve their collaboration but it degraded after a couple of minutes without feedback that they were on the right track. Additionally, the students read the warnings that appeared with each adaptation, but they complained that they felt they were already doing what was asked of them. When *voting* triggered a second time, they were quickly able to get back on track, maintain a shared focus but with little discussion or deep engagement with the content.

As a result of these observations from the pilot study, we reduced the length of time a group had to sustain high-quality collaboration in order to remove the adaptations from four minutes to two minutes. We also added an additional adaptation, *prompt* (Figure 1, top right), before *control lockout*. This adaptation simply provides students with a reminder of what is expected of them and gives them the opportunity to self-correct before the restrictive adaptations are triggered. The time that warning messages remained on-screen before activating *control lockout* and *voting* was reduced to 20 seconds as the video showed that to be enough time for the students to read the message. The rest of the implementation details remained the same.

## Data analysis

Formal analysis of all study sessions began by coding the videos for social regulation using the same codes that were used to develop the collaboration quality detection approach (Evans et al., 2016). The bulk of the coding was carried out by a doctoral student who was unfamiliar with the adaptations being evaluated. To establish inter-rater reliability, the student and the first author independently coded a session from the baseline data collection phase of the study. Codes were applied to episodes (Chi, 1997), and each episode could contain multiple codes. The majority of codes had a Cohen's kappa ( $\kappa$ ) above 0.61, typically considered "substantial" agreement, with several codes above 0.81, or "almost perfect" agreement (Landis & Koch, 1977).

To determine if the adaptations were triggered appropriately, we looked at the video codes in the

intervals leading up to the triggering of each adaptation. If the video codes that occurred in that same interval were primarily negative, the adaptations were considered appropriate. We also looked for video intervals that showed improvement while adaptations were active to determine if the improvements were detected. Finally, we looked for video intervals that showed collaboration problems that were undetected by the computer.

To understand how students responded to the adaptations, we first looked at the video codes in intervals immediately following the activation of an adaptation. For an adaptation to be considered successful, the video codes should show improved collaboration. Due to the frustration observed in the pilot, we also reviewed the videos from the evaluation sessions to understand students' emotional responses to the adaptations.

## Results

The results from the summative evaluation sessions suggest that the revised approach to triggering adaptations was more effective. Collaboration improved, along some dimensions, immediately following the first interventionist adaptation in the sequence (*prompt*) every time it was triggered, meaning that later adaptations were never triggered. This is a positive result because the *prompt* appeared to lead to more effective collaboration very quickly and consistently, but a side effect is that most of the adaptations were not therefore tested in this phase of the study. Additionally, the adaptations could not address all collaboration problems—disengaged students who showed no inclination to participate in the group work remained disengaged whether or not adaptations were present, and off-task interactions taking place away from the tabletop computer could not be detected. The biggest positive impacts of the adaptations were reduced disruption caused by individual students and less time spent engaged in low-quality collaboration.

Each group used the tabletop computer twice during the final evaluation phase of the study. Of the two groups that used the tabletop with adaptations available (Groups 1 and 3), only Group 1 triggered the adaptations. In Group 1's first session in the evaluation phase, they triggered the *prompt* once, at 4 minutes into the session. The video showed that the prompt was triggered at an appropriate time, after several minutes of the group being off task and pressing buttons on-screen without any explicit coordination. The prompt included a reminder that students should make sure they understood the task goal and that they were working toward it. When the prompt appeared, the group did revisit the instructions. The sequence of adaptations did not progress to the next stage because, immediately after reading the instructions, the group began to engage in on-task work. However, only some students in the group were engaged during most subsequent episodes. The disengaged, off-task students typically refrained from touching the computer after the prompt appeared in this session and were therefore undetectable. This behavior was noticeably different from Group 1's baseline and pilot sessions, in which these students would attempt to interact with the screen without fully engaging in the activity, disrupting students who were engaged.

In Group 1's second session in the evaluation phase, they triggered the *prompt* twice, at 9 and 17 minutes. The sequence did not progress beyond the *prompt* in either case. The video analysis showed that, in the intervals leading up to the first prompt, the whole group was engaging in primarily low-quality collaboration for around a minute, followed by a period of high-quality collaboration between two students with the other two students completely off task. After the prompt was dismissed, the off-task students continued to be off-task but refrained from touching the screen while the other two students worked collaboratively. In the intervals leading to the second instance of the *prompt*, one of the off-task students became interested in the screen, trying to take control of a particular object by repeatedly hammering on it. This interaction was highly disruptive to the engaged students, who were close to completing the task. When the prompt appeared for the second time, the engaged students appeared annoyed by it but they were able to dismiss it quickly and it deterred the other student from hammering on the screen. He sat back from the computer but made some verbal contributions to the collaboration—encouraging his teammates as they solved the assigned task.

We also compared the length of Group 1's periods of sustained low-quality collaboration taking place at the computer in three types of intervals: (1) "pre-adaptation"—intervals that cause an adaptation to trigger; (2) "adaptation active"—intervals in which an adaptation is active; and (3) "no adaptation"—intervals during which no adaptations are present. "Periods of sustained low-quality collaboration" means periods of time with one or more continuous episodes of low-quality collaboration.

Table 1 shows that the adaptations appeared to reduce the length of periods of low-quality collaboration involving the computer. The median length of periods of low-quality collaboration was consistently longer during intervals that caused an adaptation to trigger than during other intervals in all sessions where adaptations were present. In the baseline session, when the adaptations were not in use, the median length of sustained low-quality collaboration was at least twice that of the sessions where adaptations were available. In both final evaluation sessions, there were considerably more occurrences of low-quality collaboration when no adaptations were present than in other intervals. This effect occurred because, when there were no

adaptations present, occurrences of low-quality collaboration were brief and punctuated by high-quality collaboration, causing the number of occurrences to increase and the length of the occurrences to decrease.

Table 1: The median length (in seconds) of sustained periods of low-quality collaboration in Group 1's sessions

| Interval Type     | Baseline          |                  | Pilot             |                  | Evaluation 1      |                  | Evaluation 2      |                  |
|-------------------|-------------------|------------------|-------------------|------------------|-------------------|------------------|-------------------|------------------|
|                   | Median length (s) | # of occurrences | Median length (s) | # of occurrences | Median length (s) | # of occurrences | Median length (s) | # of occurrences |
| Pre-adaptation    |                   |                  | 16                | 7                | 106               | 1                | 15                | 5                |
| Adaptation active |                   |                  | 9.5               | 19               | 0                 | 0                | 11                | 2                |
| No adaptation     | 20                | 11               | 9                 | 6                | 8                 | 19               | 10                | 15               |

Although Group 3 did not trigger the adaptations, the video analysis showed a serious collaboration problem—only one student engaged with the task while the other two group members sat back from the tabletop computer, chatting and using their phones. This highlights a known limitation of the detection approach used in our system, namely that it is only able to detect interactions with the screen (Evans et al., 2016).

## Discussion and conclusion

The results show that our system was able to detect certain collaboration problems—primarily disruption caused by individual students and poor coordination among group members. Adaptations were triggered appropriately in these instances. How students in Group 1 received the adaptations differed by how motivated they were to engage with the activity—motivated students adjusted their behavior in a positive direction but already disengaged students were put off completely. Although this effect led to positive outcomes for the motivated students, who were able to make progress where they had previously been blocked by disruptive students, it was problematic for the disengaged students.

Although Group 3 never triggered the sequence of adaptations, both Group 1 and 3 saw the *group awareness icon* change color in response to the collaboration quality detected by the computer. However, with the exception of a single utterance in the pilot, there was no evidence that either group made use of the icon.

The *prompt* appeared to be effective at encouraging Group 1 to think about how they were interacting and to make sure they were working on the task as assigned. The first time the prompt appeared, the students took time to read it and follow its advice. The second and third time it appeared, the students were quicker to dismiss it, possibly due to familiarity, but both times, it caused an off-task student to stop disruptive behavior.

We consider the fact that, once *prompt* was added to the sequence of adaptations, no further adaptations were triggered, to be a positive outcome for this work. In all cases, the prompt was followed by sustained periods of high-quality collaboration, albeit only for those students who were engaged in the task. Beyond this study, the ideal outcome of using these adaptations over a longer period of time would be that they render themselves unnecessary—with such an outcome it would be possible to conclude that the adaptations successfully scaffold effective collaboration, fading once a group has adopted the principles that the adaptations support (Pea, 2004).

In this paper, we have described the design, implementation, and evaluation of a set of tabletop software adaptations to encourage effective collaboration when problems are detected. Presented as a sequence that grows increasingly restrictive if collaboration does not improve, our adaptations showed promise as supports for groups that struggle with disruptive behavior. Due to the small scale of our classroom field evaluation, further study is needed to determine the extent of our approach's effectiveness, but overall, our work demonstrates that tabletop applications that can detect and adapt to poor-quality collaboration can encourage more effective group work by deterring disruptive behavior.

## References

- Chi, M. T. H. (1997). Quantifying Qualitative Analyses of Verbal Data: A Practical Guide. *Journal of the Learning Sciences*, 6(3), 271–315.
- Dillenbourg, P., & Jermann, P. (2007). Designing Integrative Scripts. In F. Fischer, I. Kollar, H. Mandl, & J. M. Haake (Eds.), *Scripting Computer-Supported Collaborative Learning: Cognitive, Computational and Educational Perspectives*. New York: Springer.
- Evans, A. C., Davis, K., Fogarty, J., & Wobbrock, J. O. (2017). Group Touch : Distinguishing Tabletop Users in Group Settings via Statistical Modeling of Touch Pairs. In *Proceedings of CHI 2017*, 35–47. New

- York: ACM Press.
- Evans, A. C., Wobbrock, J. O., & Davis, K. (2016). Modelling collaboration patterns on an interactive tabletop in a classroom setting. In *Proceedings of CSCW 2016*, 860–871. New York: ACM Press.
- Evans, A., & Wobbrock, J. O. (2014). Filling in the gaps: capturing social regulation in an interactive tabletop learning environment. In *Proceedings of ICLS*, 1157–1161. Boulder, Colorado: ISLS.
- Facing the Future, & Snow Leopard Trust. (2009). *Engaging Students in Conservation: Protecting the Endangered Snow Leopard*. Seattle, WA. Retrieved from <http://www.snowleopard.org/learn/resources-for-educators>
- Hadwin, A. F., Oshige, M., Gress, C. L. Z., & Winne, P. H. (2010). Innovative ways for using gStudy to orchestrate and research social aspects of self-regulated learning. *Computers in Human Behavior*, 26(5), 794–805.
- Järvelä, S., & Hadwin, A. F. (2013). New Frontiers: Regulating Learning in CSCL. *Educational Psychologist*, 48(1), 25–39.
- Järvelä, S., Kirschner, P. A., Panadero, E., Malmberg, J., Phielix, C., Jaspers, J., Koivuniemi, M., & Järvenoja, H. (2014). Enhancing socially shared regulation in collaborative learning groups: designing for CSCL regulation tools. *Educational Technology Research and Development*, 63(1), 125–142.
- Kumar, R., & Rosé, C. P. (2011). Architecture for Building Conversational Agents that Support Collaborative Learning. *IEEE Transactions on Learning Technologies*, 4(1), 21–34.
- Landis, J., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1).
- Malmberg, J., Järvelä, S., Järvenoja, H., & Panadero, E. (2015). Computers in Human Behavior Promoting socially shared regulation of learning in CSCL : Progress of socially shared regulation among high- and low-performing groups. *Computers in Human Behavior*, 52, 562–572.
- Martinez-maldonado, R., Kay, J., Yacef, K., Edbauer, M. T., & Dimitriadis, Y. (2013). MTClassroom and MTDashboard: supporting analysis of teacher attention in an orchestrated multi-tabletop classroom. In *Proceedings of CSCL*, 320–327. Madison, WI: ISLS.
- Martinez-Maldonado, R., Yacef, K., Kay, J., Al-Qaraghuli, A., & Kharrufa, A. (2011). Analysing frequent sequential patterns of collaborative learning activity around an interactive tabletop. In *Proceedings of EDM*, 111–120. International Educational Data Mining Society.
- Martinez, R., Collins, A., Kay, J., & Yacef, K. (2011). Who did what? Who said that?: Collaid: an environment for capturing traces of collaborative learning at the tabletop. In *Proceedings of ITS*, 172–181. New York: ACM Press.
- Panadero, E., & Järvelä, S. (2015). Socially Shared Regulation of Learning: A Review. *European Psychologist*, 20(3), 190–203.
- Pea, R. D. (2004). The social and technological dimensions of scaffolding and related theoretical concepts for learning, education, and human activity. *Journal of the Learning Sciences*, 13(3), 423–451.
- Rogat, T. K. & Linnenbrink-Garcia, L. (2011). Socially Shared Regulation in Collaborative Groups: An Analysis of the Interplay between Quality of Social Regulation and Group Processes. *Cognition and Instruction*, 29(4), 375–415.
- Trausan-Matu, S., Dascalu, M., & Rebedea, T. (2014). PolyCAFe-automatic support for the polyphonic analysis of CSCL chats. *International Journal of Computer-Supported Collaborative Learning*, 9(2), 127–156.
- Volet, S., Vauras, M., & Salonen, P. (2009). Self- and social regulation in learning contexts: An integrative perspective. *Educational Psychologist*, 44(4), 215–226.
- Wang, H. C., Rosé, C. P., & Chang, C. Y. (2011). Agent-based dynamic support for learning from collaborative brainstorming in scientific inquiry. *International Journal of Computer-Supported Collaborative Learning*, 6(3), 371–395.
- Wise, A. F. (2014). Designing pedagogical interventions to support student use of learning analytics. In *Proceedings of LAK*, 203–211.

## Acknowledgments

We thank Cindy Hmelo-Silver for her guidance on this project. This work was supported in part by the National Science Foundation under awards IIS-0952786 and IIS 1053868. Any opinions, findings, conclusions or recommendations expressed in this work are those of the authors and do not necessarily reflect those of the National Science Foundation.

# Integrative Visualization: Exploring Data Collected in Collaborative Learning Contexts

Ben Rydal Shapiro, Georgia Institute of Technology, ben@benrydal.com

**Abstract:** The development and use of computational approaches to make sense of data collected in collaborative learning contexts is expanding rapidly in the Computer-Supported Collaborative Learning (CSCL) and Learning Sciences (LS) communities. However, developing and using these approaches in ways that maintain a commitment to theory and situational context is a significant challenge. This paper proposes *integrative visualization*, a computationally assisted, human-centered process to visually explore data collected in collaborative learning contexts in a way that maintains a commitment to theory and situational context. To do so, this paper summarizes how integrative visualization supported the development and use of interaction geography, an approach to describing, representing, and interpreting collaborative interaction across the physical environment. This paper concludes by emphasizing the need to further develop integrative visualization in the CSCL and LS communities through stronger connections with the information visualization community.

## Introduction

The development and use of computational approaches to make sense of data collected in collaborative learning contexts is expanding rapidly in the Computer-Supported Collaborative Learning (CSCL) and Learning Sciences (LS) communities. In particular, computational approaches are providing ways to extend traditional data analysis of phenomena such as language, gesture, or gaze important to studying collaborative learning (Rosé et al., 2008; Schneider & Pea, 2014). Alternatively, these approaches offer resources to reason about larger data sets and data scales including multimodal data collected across classroom settings (Knight et al., 2017). However, developing and using these approaches in ways that maintain a commitment to theory and situational context is a significant challenge (Wise & Schwarz, 2017; also see Law et al., 2018). This challenge also reflects broader efforts in and outside of education to develop “orientations towards and frameworks for data science that are not reducible to mere efficiency” (Zegura, DiSalvo & Meng, 2018; also see Kahn, 2017; Berland et al., 2017).

This paper proposes *integrative visualization*, a computationally assisted, human-centered process to visually explore data collected in collaborative learning contexts in a way that maintains a commitment to theory and situational context. Integrative visualization extends visual methods of exploratory data analysis (Tukey, 1977) to iteratively transcribe data collected in context in order to organize data and develop codes, categories, units of analysis, and questions that support the development of grounded theory (Glaser & Strauss, 1967) and new computational tools. In particular, integrative visualization extends the practice of “integrative diagramming” (Strauss, 1987) through techniques of exploratory data analysis and is summarized as follows:

data context → transcription amplified by visualization → grounded theory & computational tools

This paper illustrates this process of integrative visualization by describing how it supported the development and use of interaction geography, an approach to describing, representing, and interpreting collaborative interaction across the physical environment (Shapiro, Hall & Owens, 2017; Shapiro & Hall, 2018). This paper concludes by emphasizing the need to further develop integrative visualization in the CSCL and LS communities through stronger connections with the information visualization community.

## An example of integrative visualization

### Data context

The example described in this paper draws from data collected during a three-year project in collaboration with a nationally renowned museum located in the mid-South region of the United States. This project sought to understand how visitors to this museum cultivated interests in and learned about the diverse historical and cultural heritage of American Roots music during and after their visit. Two initial research questions guided this work. First, this work aimed to understand the organization of visitors’ activity not only at single museum exhibits as was typically the case in existing research but also as visitors moved across gallery spaces over their complete museum visit. Second, this work sought to understand how visitors furthered their own interest-driven engagement and learning through their activity across gallery spaces.



To answer these questions, a purposive sample of complete museum visits across 22 visitor group cases (2–5 visitors per group) was collected over a period of six weeks. This sample was collected in close collaboration with museum partners, many generous families and visitors participating in this research, and followed institutional review board (IRB) protocols. Data from these 22 cases included continuous, multi-perspective video and audio records (72 hrs total) of each visiting groups’ movement, interaction, and social media use collected through small, unobtrusive cameras that each visitor in a group wore as necklaces for the duration of their visit with no researchers present (visits ranged from 30 min to 4 hrs). This data provided detailed records of visitors’ interaction (e.g., conversation, movement, use of cell phones and cameras).

This data presented two novel research challenges. First, existing research was only beginning to make sense of complex, multi-perspective audio and video data in ways necessary to answer the questions informing this work (see Marin, 2013; Steier, 2014; Taylor & Hall, 2013 for early work). Second, apart from a few novel and inspirational examples (see Fouse et al., 2011), computational tools to support this work did not exist. For example, contemporary transcription software (e.g., InqScribe, NVivo) and video editors (e.g., Final Cut Pro, Adobe Premier) were not designed to account for the spatial dimension of collaborative interaction as people moved. Likewise, information visualization and visual analytics software used to study visitor activity in settings such as museums did not operate at a fine enough grain size to answer the research questions informing this work and also typically focused exclusively on visitors’ movement ignoring their conversation.

In summary, this particular cultural heritage museum along with these initial questions, goals, types of data collected, and unique challenges framed the data context of this work. This data context included a theoretical commitment to studying visitors’ interest-driven engagement and learning as they moved across gallery spaces through continuous, multi-perspective video records of visitor activity. The following section uses a set of figures and analysis to illustrate an iterative process of transcribing this data amplified by visualization in a way that maintains this theoretical commitment and is driven by this data context.

### Transcription amplified by visualization

Figure 1 is a hand drawn sketch that represents a first attempt to make sense of this data and specifically, a five-member family’s activity within a particular museum gallery space. Namely, the sketch tries to make sense of the family’s experiences across a gallery space from watching five separate video records.

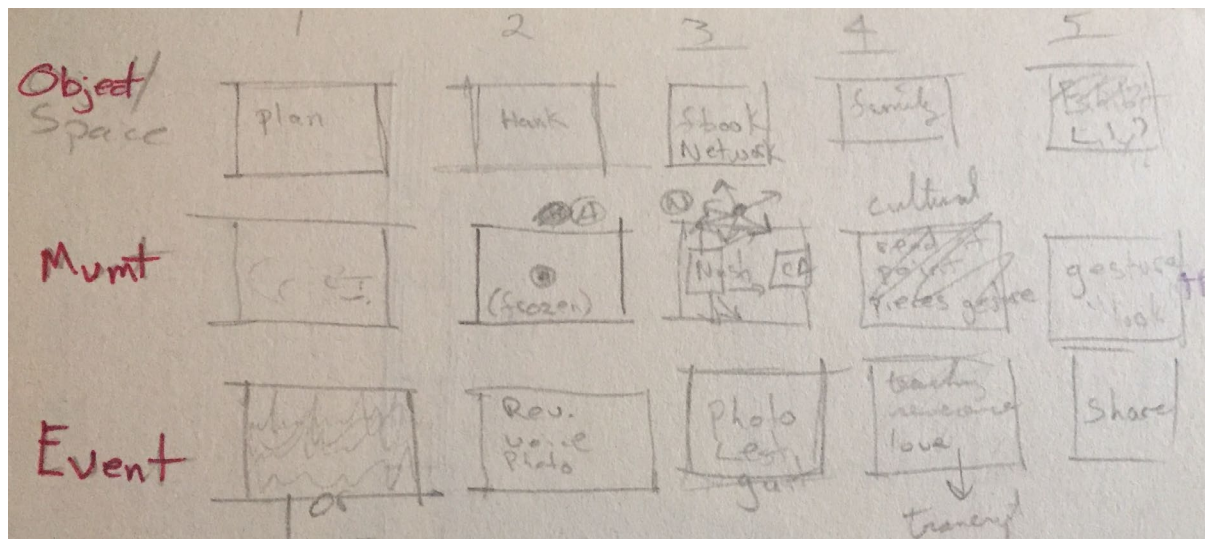


Figure 1. Manual sketch of a family’s activity in a museum gallery transcribed from multi-perspective video.

Importantly, the sketch draws from a set of famous architectural drawings known as the “Manhattan Transcripts” developed by architect Bernard Tschumi: The Manhattan Transcripts aimed to expand the theorization and design of architecture beyond space and form to also include the relations between space, movement, and event or what happens in space (Tschumi, 1994). The sketch uses the Manhattan Transcripts to organize aspects of this family’s activity in this gallery space across dimensions of object/space, movement, and event. For example, the 2<sup>nd</sup> column describes one family member’s (named Adhir) activity at an exhibit. The first row of this column describes the exhibit/object of focus (a musician named Hank Williams); the second row characterizes Adhir’s movement as frozen or not moving; the third row characterizes events/things Adhir does at the exhibit (stands in reverence, talks, takes a photograph). Altogether, the figure shows a commitment to



understanding visitor groups as a mobile unit (e.g., multiple participants are placed in columns across different times of their visit) and questions concerning how and what happens as each family member moves.

Figure 2 is a computationally generated sketch produced 4 months after the previous sketch. It illustrates an early attempt to transcribe the same five-member family's conversation as they move across part of the museum gallery space described previously. Put differently, this sketch illustrates an initial attempt to address a fundamental problem regarding how to make sense of people's conversation as they move.

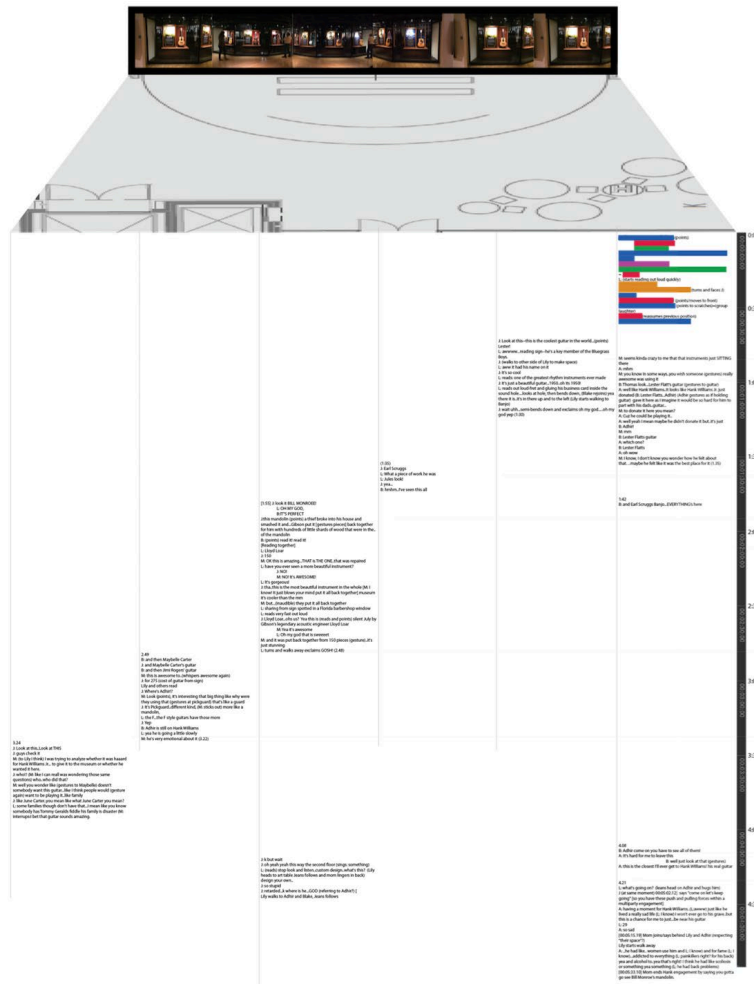
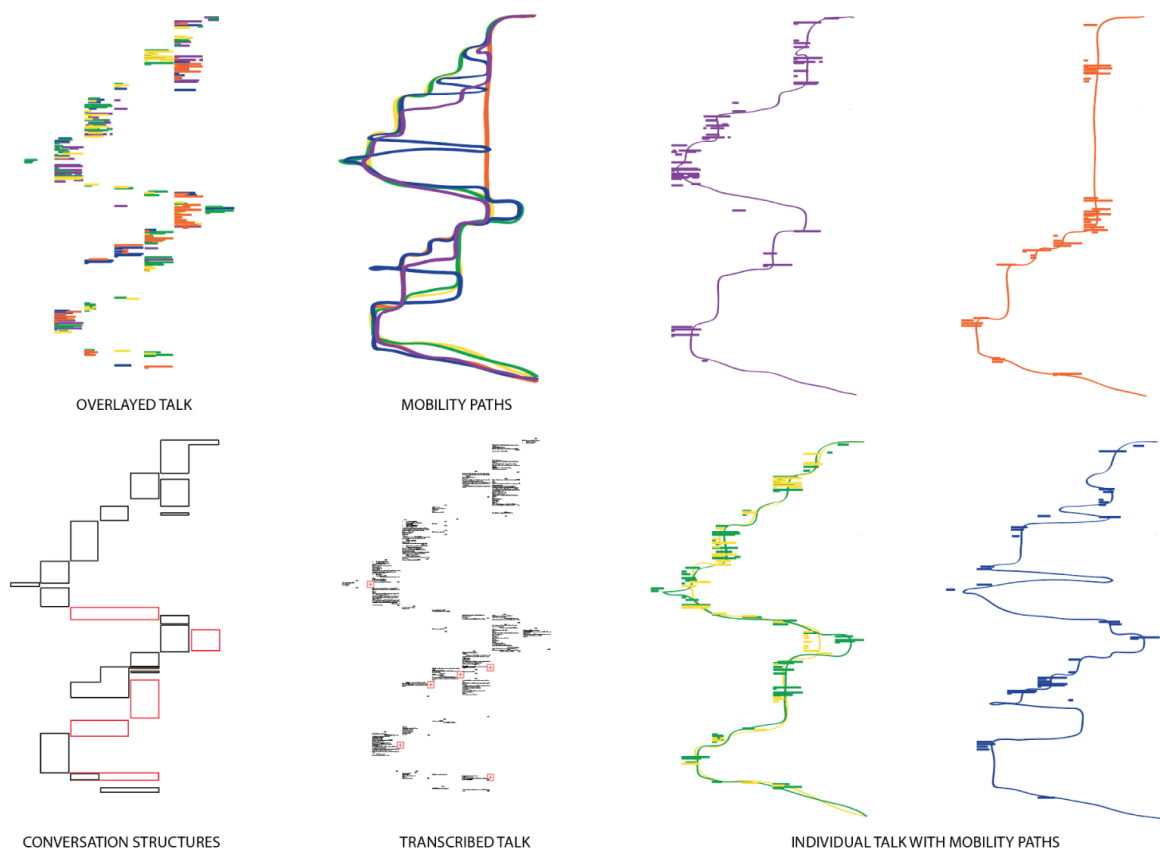


Figure 2. Extending the space-time cube to transcribe a family's conversation across a museum gallery space.

The sketch shows a floor plan of part of the gallery space (i.e., looking down on the space). This floor plan is tilted in 3D isometric perspective. A set of images is shown on top of the floor plan. These images correspond to six museum exhibits that line a semicircle drawn on the floor plan. A timeline in minutes and seconds along with grid lines that separate 6 regions of the floor plan (corresponding with the 6 exhibits along the semicircle) extend downward/below the floor plan. Turns of transcribed talk from each family member are placed along this timeline and in regions of this gridded space (i.e., placed in time and space). Turns of talk are also grouped into topically related conversations (typically spatially related conversations about exhibit content). Thus, the position of turns of talk/conversations vertically on the timeline and horizontally across the floor plan indicate when (over approximately 5 minutes) and where (along the semicircle of six exhibits) each turn of talk/conversation occurs. In addition, turns of talk in the first conversation (top of the timeline) are overlaid by colored lines that indicate speaker (i.e., color indicates which family member speaks that conversation turn).

Altogether, this sketch illustrates an initial effort to extend a geographical perspective called time geography and a visualization system known as the space-time cube (Hagerstrand, 1970) to begin to develop codes and categories to explore this family's conversation over space and time in new ways. Doing so necessitated categorizing and grouping conversation in ways that foregrounded the spatial and mobile dimensions of visitors' conversation (i.e., reflective of the theoretical commitments and data context of this work).

Figure 3, produced approximately three months after the previous figure visualizes different aspects of the same family’s interaction over the total time they spent in the same part of the gallery space (approximately 8 minutes) in a small multiple format (Tufte, 1990). The top left visualization in the figure titled “overlaid talk” shows the turns of talk from the family colored by speaker. The visualization below this titled “conversation structures” isolates visual boxes that group turns of talk into topically related conversations while the visualization adjacent titled “transcribed talk” shows all transcribed talk for the family. In comparison to Figure 2, this includes about 3 more minutes of talk. The visualization above titled “mobility paths” shows the family’s movement across this space and over time as lines or paths (color again indicates family member). The four visualizations to the right isolate each family member’s individual movement and turns of talk: the mother named Mae is shown in purple, her son and daughter named Jeans and Lily are shown together in green and yellow respectively as their movement is nearly identical, their 6-year old brother named Blake is shown in blue, and Lily’s fiancé Adhir is shown in orange. By visualizing each family member’s movement and conversation as layers over space and time comparisons can be drawn across family members to simultaneously study who speaks, what is said, and where each family member goes in this gallery space. For example, the figure supports asking new questions such as how parents’ conversation across gallery spaces structures and responds to children’s movement and conversation.



**Figure 3.** Small multiple of a family’s movement and conversation over space and time in a museum gallery.

Figure 4 presents a more refined computationally generated visualization produced four months later of two members of this family in this gallery space, Blake and Adhir (shown once again in blue and orange respectively). On the left, the floor plan is now shown in 2D (looking directly down on the space) and depicts the full gallery space. The semicircle set of six exhibits is now clearly shown and one exhibit is marked with “Williams” to indicate it features content about a famous musician named Hank Williams. Blake and Adhir’s movement is shown across the floor plan or in “floor plan view” indicating where they travel while visiting this gallery space. A timeline extends to the right of the floor plan as opposed to below/under the floor plan. In comparison to Figures 2 and 3, this subtle rotation of the timeline (suggested by a collaborator, Lara Heiberger, in the context where this data was collected) aids in interpreting complex space-time visualizations. In this case, similar to the previous figures but more clearly illustrated here, the “space-time view” (Hagerstrand, 1970) extends

Blake and Adhir’s movement on the floor plan horizontally over time. This view shows how they interact with exhibits and one another over time. For example, the space-time view shows that after entering the gallery space (top left of the floor plan view and beginning of the space-time view), Adhir and Blake walk together toward the Hank Williams exhibit. Subsequently, Adhir stands for almost 5 minutes at the Hank Williams exhibit, as indicated by his horizontal orange path in the space-time view that extends from approximately minutes 0–5 and corresponds to the vertical position of the Hank Williams exhibit in the floor plan view. In the meantime, while Adhir is standing, Blake is moving quickly (apparently running) back and forth across the gallery space (i.e., across the semi-circle of exhibits on the floor plan) in multiple attempts to draw Adhir away from the Hank Williams exhibit. After four failed attempts, Blake finally succeeds in leading Adhir on what is described as a tour of other exhibits in the gallery, indicated by their intertwined paths from approximately minutes 5-6. The change in line pattern in Blake’s path distinguishes between three different horizontal areas of space on the floor plan providing some description of horizontal movement on the floor plan in the space-time view.

In comparison to previous figures, this figure begins to provide more detailed ways to interpret collaborative interaction across a physical environment (i.e., reflective of the original commitments of this work). For instance, the figure characterizes new, path-based units of collaborative interaction such as Blake’s tour that support asking and answering new types of questions including how young children use their movement to manage their families as resources for their own interest-driven engagement and learning.

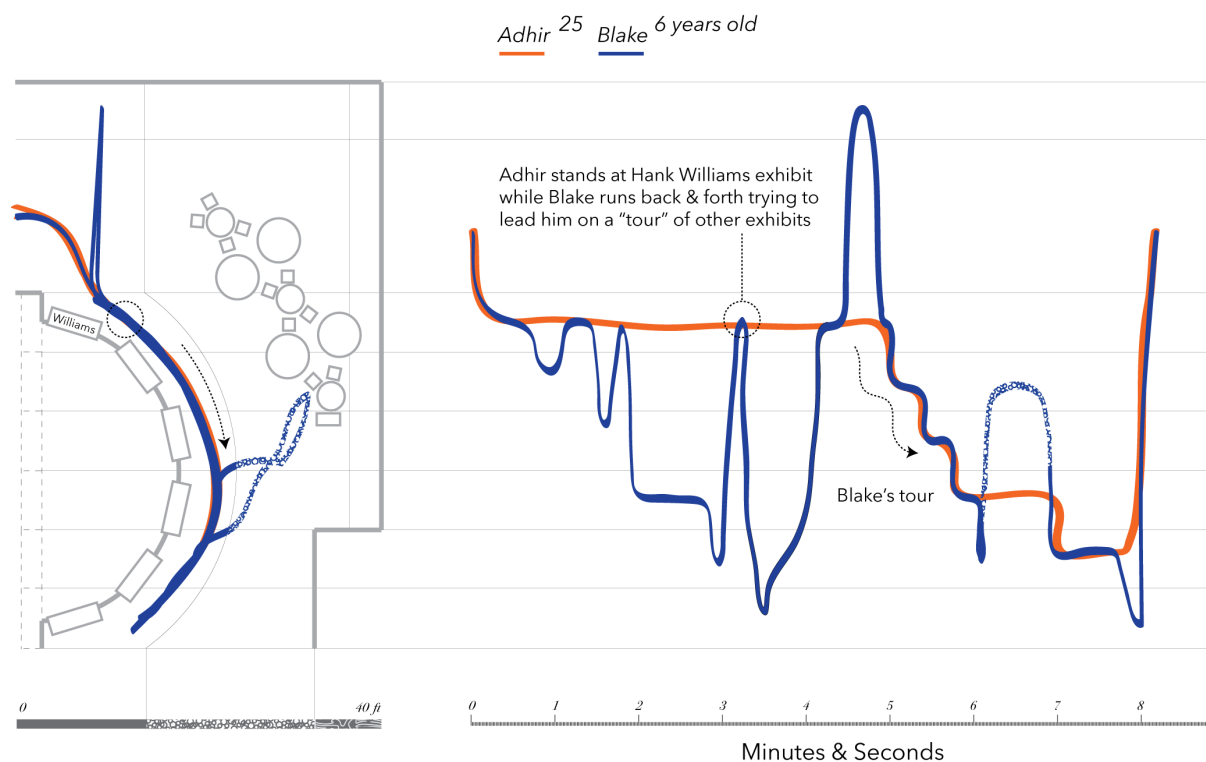


Figure 4. Adhir (orange) and Blake’s (blue) movement over space and space-time in a museum gallery.

### Grounded theory and computational tools

Figure 5 was produced over a year after Figure 4 and represents a significant amount of theoretical and computational development. The figure is a screenshot from a dynamic visualization tool called the Interaction Geography Slicer (IGS), which allows for new forms of interaction and multimodal analysis. The figure also illustrates a refined method to transcribe movement and conversation, in this case, of all five members of the previous family (including Blake and Adhir) in the previously described gallery space. This method called Mondrian Transcription draws inspiration from the Manhattan Transcripts and the Modernist artist, Piet Mondrian (1872–1944), particularly his use of lines in relation to forms, which resemble how movement and conversation are represented, coded, and categorized in Mondrian Transcription. The top half of the figure shows the family’s movement and the bottom half shows their conversation in relation to their movement (i.e., the family’s movement is shown in gray beneath their conversation to link the two halves of the figure). Conversation is transcribed and organized in ways introduced previously. First, each turn at talk is shown as a colored line to indicate which family

member speaks that conversation turn (indentations indicate overlapping speech). Second, colored lines of talk are gathered into boxes that group topically related sequences of conversation turns and movement (e.g., usually related to artifacts/musicians in this setting). In the space-time view, each box marks the start, duration, and end of a sequence. In the floor plan view, conversation turns and separate (in time) sequences accumulate within regions of gridded space--the box thickness in the floor plan view increases with each repeated sequence within a region of space (resembling a “heat map” of talk in place). For example, the region of space around the Hank Williams exhibit has the largest number of conversation turns (indicated by the many colored lines of talk) and is enclosed by a dense box that reflects five separate (in time) sequences occurring at the Hank Williams exhibit.

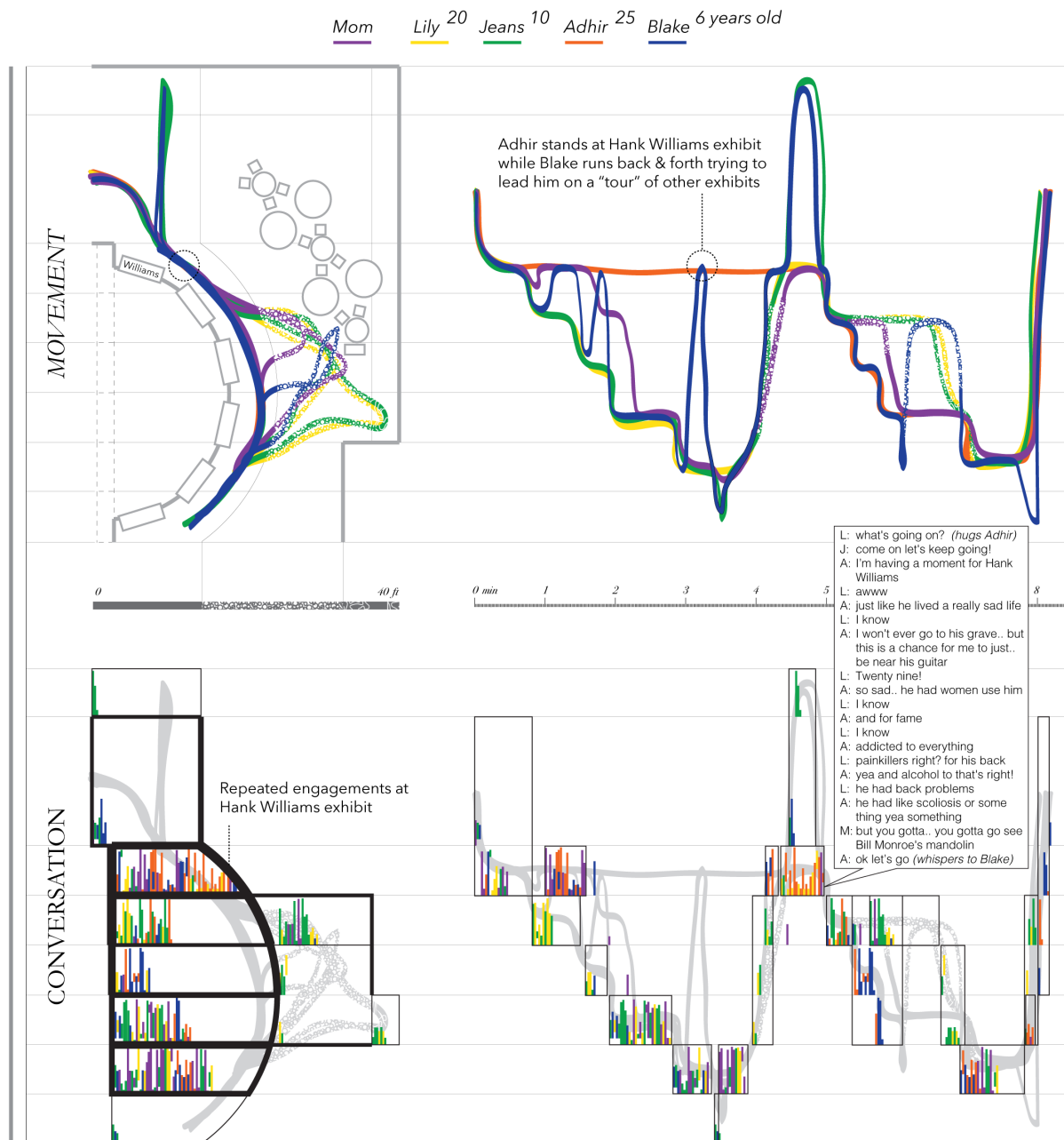


Figure 5. Screenshot from the Interaction Geography Slicer (IGS) of family’s movement and conversation.

In the figure, the highlighted sequence (i.e., readable conversation) in the space-time view expands the conversation turns of one particular sequence. In other words, the highlighted sequence illustrates an analytic “operation” possible within the IGS on data collected in this work. As the figure shows, one can use the IGS to select, magnify, visualize, and read conversation turns. Not shown is the additional ability to use the IGS to watch and listen to video/audio from the perspective of each family member gathered as part of this work. The IGS syncs

multi-perspective audio and video to visualizations such as Figure 5. As a result, anywhere a user clicks on the visualization activates audio or video from the perspective of an individual.

The figure shows how these refined tools/methods provide ways to engage in exploratory data analysis (Tukey, 1977) that pays careful, contextual attention to how people engage and learn across museum gallery spaces. For example, the highlighted conversation or sequence in the figure from approximately minutes 4–5 in the space-time view encompasses a complex mesh of activity around the Hank Williams exhibit. Reading this sequence of activity in relation to the rest of the figure shows how: 1) Lily (yellow) soothes the emotions of Adhir (her fiancé) by hugging and consoling him as he compares the Hank Williams exhibit to a “grave” (in line 8); 2) Jeans (green) gives Lily and Adhir privacy by leading a frustrated Blake away from the Hank Williams exhibit (the extension of their movement paths upwards in the floor plan and space-time views indicating their movement away from the exhibit); 3) Blake and Jeans rejoin Lily and Adhir as Adhir continues to share his own account of Hank William’s painful life; 4) Mae (Mom in purple), who has been standing near Adhir and Lily and observing her family’s interaction, helps Blake lead Adhir on a tour of other exhibits by saying to Adhir, “but you gotta.. you gotta go see Bill Monroe’s mandolin” (in lines 22–23); and 5) Evidently fully aware of Blake’s ongoing project to lead a tour, Adhir whispers to Blake, “ok let’s go” and they move forward together to the next exhibit along the semicircle of exhibits (at the end of the highlighted conversation; see Shapiro, Hall & Owens, 2017).

Altogether, the ability to read this sequence of activity in relation to the rest of the figure by exploring data within the IGS reveal phenomena and relations between phenomena such as Blake’s tour, Adhir’s persistent engagement with the Hank Williams exhibit, a mother’s efforts to support her children’s engagement, and particularly important units of engagement or “peak engagement contours” such as the previously described sequence of activity during this family’s visit to this gallery space. In other words, the figure communicates concepts and methods of interaction geography, a new approach to describing, representing, and interpreting people’s collaborative interaction as they move across the physical environment, and how to use interaction geography to study visitors’ interest-driven engagement and learning across a museum gallery space.

## Discussion and conclusion

In summary, the development and use of interaction geography was interleaved with and furthered a line of exploratory data analysis that led to new questions, units of analysis, and computational tools to describe how visitors pursued their own interest-driven engagement and learning in a museum. Put differently, the example characterizes integrative visualization, a computationally assisted, human-centered process to visually explore data in a way that maintains a commitment to theory and situational context.

Though illustrated through a single example in this paper, integrative visualization is proposed as a generalizable process researchers or practitioners can leverage to make sense of data collected in collaborative learning contexts. Integrative visualization begins with a data context, understood as an initial set of questions, goals, challenges, and particular types of data collected in a specific context. In the example, this data context included continuous, multi-perspective audio/video data collected in a cultural heritage museum and a theoretical commitment to understanding visitors’ interest-driven engagement and learning as they moved across gallery spaces. Subsequently, integrative visualization extends visual methods of exploratory data analysis to iteratively transcribe data. Like any process of transcription, transcription of data amplified by visualization is theory laden and selective because it aims to organize data and develop codes, categories, questions, and units of analysis central to the data context (see Ochs, 1979; Hall, 2000). For instance, the example illustrated the use of the space-time cube to categorize conversation over space and time in ways that foregrounded mobility. Finally, integrative visualization supports the development of grounded theory and new computational tools. On one hand, the example characterized concepts and methods of interaction geography, units of analysis, and questions about visitors’ interest driven engagement and learning as potentially generalizable to other types of settings where people move to engage and learn (e.g., other museums, natural or urban environments). On the other hand, the example demonstrated how integrative visualization produced computational tools that support new forms of interaction and multimodal analysis. Importantly, developing these tools required (and requires) shifting from using traditional transcription tools (e.g., ranging from Microsoft Word to InqScribe to NVivo) to composing in more dynamic graphical layout tools (e.g., Adobe products) to developing visualizations and software in programming languages used in exploratory data analysis such as Processing and p5.js (see Fry, 2004). Though fully describing this shift is beyond the scope of this paper, it is critical to integrative visualization.

Integrative visualization may be particularly well suited to addressing some of the unique challenges of making sense of unstructured data (e.g., audio and video) about people’s interaction collected in and important to studying collaborative learning contexts. However, such work entails stronger connections between the CSCL and LS communities and the information visualization community. These connections are challenging to develop but essential to integrating technical skills of parsing, mining, representing, and interacting with data (see Stasko



et al., 2008; Fry, 2004) with non-technical ways of working with data that foreground theoretical and contextual dimensions (see Wise & Schwarz, 2017; Kahn, 2017; Zegura, DiSalvo & Meng, 2018; DiSalvo, 2016).

## References

- Berland, M., Halverson, E., Polman, J. & Wilkerson, M. (2017). Expressive construction: Enabling learners to represent powerful ideas. In J. Roschelle, W. Martin, J. Ahn, & P. Schank (Eds.), *Cyberlearning Community Report: The State of Cyberlearning and the Future of Learning with Technology*.
- DiSalvo, B. 2016. Participatory Design through a Learning Science Lens. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 4459-4463.
- Fouse, A., Weibel, N., Hutchins, E., and Hollan, J.D. (2011). ChronoViz: a system for supporting navigation of time-coded data. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems (CHI EA '11)*. ACM, New York, NY, USA, 299-304.
- Fry, B. J. (2004). Computational information design. Ph.D. Dissertation. Cambridge: MIT.
- Glaser, B.G. & Strauss, A.L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago, IL: Aldine.
- Hagerstrand, T. (1970). What about people in regional science? *Papers in regional science*, 24(1), 7-24.
- Hall, R. (2000). Video recording as theory. In D. Lesh & A. Kelley (Eds.) *Handbook of Research Design in Mathematics and Science Education* (pp. 647-664). Mahweh, NJ: Lawrence Erlbaum.
- Kahn, J. B. (2017). At the intersection of self and society: Learning, storytelling, and modeling with big data. Ph.D. Dissertation. Vanderbilt University.
- Knight, S., Wise, A., Ochoa, X. & Hershkovitz, A. (2017). Learning analytics: Looking to the future. *Journal of Learning Analytics*, 4(2), 1-5.
- Law, N., Rose, C.P., Cress, U. and Ludvigsen, S. (2018). Different technologies, methodologies and epistemologies—is CSCL a community or communities? *International Journal of Computer-Supported Collaborative Learning*, 13(2), 131-136.
- Marin, Ananda M. (2013) *Learning to Attend and Observe: Parent-child Meaning Making in the Natural World*. Ph.D. Dissertation. Northwestern University.
- Ochs, E. (1979). Transcription as theory. In E. Ochs & B. Schieffelin (Eds.), *Developmental pragmatics* (pp. 43-72). New York: Academic Press.
- Rosé, C., et al. (2008). Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *International Journal of Computer-Supported Collaborative Learning*, 3(3), 237-271.
- Schneider, B., & Pea, R. (2014). Toward collaboration sensing. *International Journal of Computer-Supported Collaborative Learning*, 9(4), 371-395.
- Shapiro, B.R., Hall, R. & Owens, D. (2017). Developing & Using Interaction Geography in a Museum. *International Journal of Computer-Supported Collaborative Learning*, 12(4), 377-399.
- Shapiro, B.R. and Hall, R. (2018). Personal Curation in a Museum. In *Proceedings of the ACM in Human-Computer Interaction*, Vol. 2, CSCW, Article 158 (November 2018). ACM, New York, NY.
- Stasko, J., Görg, C., & Liu, Z. (2008). Jigsaw: Supporting Investigative Analysis through Interactive Visualization. *Information Visualization*, 7(2), 118-132.
- Steier, R. (2014). Posing the question: Visitor posing as embodied interpretation in an art museum. *Mind, Culture, and Activity*, 21(2), 148-170.
- Straus, A.L. (1987). Integrative diagrams and integrative sessions. In A.L. Strauss, (Ed.), *Qualitative Analysis for Social Scientists* (pp. 170-182). Cambridge University Press.
- Taylor, K. H., & Hall, R. (2013). Counter-mapping the neighborhood on bicycles: Mobilizing youth to reimagine the city. *Technology, Knowledge and Learning*, 18(1-2), 65-93.
- Tschumi, B. (1994). *The Manhattan Transcripts*. Wiley.
- Tukey, John W (1977). *Exploratory Data Analysis*. Addison-Wesley.
- Tufte, E. R. (1990). *Envisioning Information*. Cheshire, Connecticut: Graphic Press.
- Wise, A., & Schwarz, B. (2017). Visions of CSCL: Eight provocations for the future of the field. *International Journal of Computer-Supported Collaborative Learning*, 12, 423-467.
- Zegura, E., C. DiSalvo, & A. Meng. (2018). Care and the Practice of Data Science for Social Good. In *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*. ACM, New York, NY.

## Acknowledgements

Amanda Meng, Jennifer Kahn, and Rogers Hall provided critically important feedback that contributed to the development of ideas in this paper. Figures copyright © Ben Rydal Shapiro. Reprinted with Permission.

# An Empirical Study of Educational Robotics as Tools for Group Metacognition and Collaborative Knowledge Construction

Chrysanthos Socratous, Cyprus Interaction Lab, Cyprus University of Technology,  
chrysanthos@cyprusinteractionlab.com

Andri Ioannou, Cyprus Interaction Lab, Cyprus University of Technology, Research Center on Interactive Media, Smart Systems and Emerging Technologies (RISE), andri@cyprusinteractionlab.com

**Abstract:** The affordances of Educational Robotics (ER) for advancing teaching and learning has become a widely researched topic. This study aims to identify the major components of collaborative knowledge construction in an ER learning environment and to investigate the mediating role of ER as mindtools to support group metacognition. Data analysis involved a micro-level examination of students' discourse, interaction with the technology, peers and the facilitator, using fine-grained analysis of video and audio recordings. The results made evident that metacognition, along with questioning and answering, were prevalent elements of collaborative knowledge construction around ER. We support that ER can be used as a learning tool and can be effective in supporting group metacognition through immediate feedback, openly accessible programmability and students' embodied interaction with the physical robot. Beyond the instrumental role of ER for supporting metacognitive processes in CSCL settings, the study provides initial evidence for a temporal relation of metacognitive talk to collaborative talk in group problem-solving.

## Introduction

The use of Educational robotics (ER) in educational contexts to support teaching and learning has become an extensively researched topic. ER is constructible and programmable high-tech devices which can be employed in education as constructivist learning tools to support teaching and learning through hands-on activities. ER too early earned an influential role as a research field, motivating the attention of many schools, and universities, both from an instructive and a research point of view. Jonassen (2000) first introduced the theoretical background and the motivation for the integration of robotic technologies as cognitive tools which can improve and enrich the educational process. According to Gaudiello and Zibetti (2013), two features of ER are linked to their high educational potential; "transparency" and "interactivity." "Transparency" refers to the openly accessible programmability of the robot whilst, "interactivity" refers to the immediacy of the feedback given by the robot when a student programs and executes the commands (Gaudiello & Zibetti, 2013).

Many studies have focused on exploring the affordances of ER in promoting several transversal skills such as problem-solving (e.g., Atmatzidou, Demetriadis & Nika, 2018), collaboration (e.g., Ardito, Mosley & Scollins, 2014), and computational thinking (e.g., Bers, Flannery, Kazakoff & Sullivan, 2014; Constantinou & Ioannou, 2018). Still, ER as metacognitive tools have been considered only recently (e.g., La Paglia, Caci, La Barbera & Cardaci, 2010; Gaudiello & Zibetti, 2013) and the research evidence is inconsistent. Further investigation in the area is needed to fully understand the potential of ER in supporting students' metacognitive processes and especially socially-mediated metacognitive processes in CSCL settings.

We present an empirical investigation of ER in a CSCL experience aiming at engaging students in collaboration and co-construction of shared understandings in the mathematics domain. In this work, the whole experience around using the robot is seen as a metacognitive experience that assists students to become more aware of their process of thinking and learning. We aim to unfold the elements of collaborative knowledge construction, identify details of the metacognitive processes during students' interaction with the robot and their peers, and document the educational potential of ER as tools for supporting group metacognition. Specifically, the research questions of this study are the following:

1. What are the elements of collaborative knowledge construction in an ER learning environment?
2. How does ER help to activate group metacognitive processes?
3. What is the relationship between collaborative talk and metacognitive talk in ER learning settings?

In the following lines, we present the theoretical framework of the study, findings from previous empirical studies, methodology, and findings from the present investigation along with discussion of the implications of this work.

## Theoretical framing

## Metacognition

Whilst, over the past years various theoretical models of metacognition have evolved, researchers agree that metacognition consists of three or at least two fundamental processes. According to Schraw and Moshman (1995), metacognition can be divided into a knowledge component (knowledge of cognition) and a skill component (regulation of cognition). Other researchers have expanded this model suggesting a three-tier model of metacognition namely, metacognitive knowledge, metacognitive judgments and monitoring, and self-regulation and control (Pintrich, Wolters, & Baxter, 2000). More recently Efklides (2011) proposed a complex interplay between the task level, the person level (where metacognitive knowledge and skills are located) and the interaction level (where metacognitive experiences take place).

There are already some studies examining the use of ER to promote metacognition at the individual level. For example, a study by Keren and Fridin (2014), examined how ER can assist the teaching of geometric thinking and promote children's metacognitive development. Findings from the study showed that students' performance on geometric thinking and metacognitive tasks were improved because of their participation in ER activities. Also, to investigate the process of constructing and programming robots as metacognitive tools, La Paglia et al. (2010) found that ER may be conceptualized as a novel metacognitive setting that motivates learners to monitor and control their own learning actions. Gaudiello and Zibetti (2013) tried to identify and classify the heuristics that are applied by elementary school students while they interact with and control ER technologies. The results demonstrated three main types of heuristics: (a) procedural-oriented, (b) declarative-oriented, and (c) metacognitive-oriented. Atmatzidou, Demetriadis, and Nika (2018) investigated the development of students' metacognitive skills in ER activities when the facilitators performed different levels of guidance (low and high) in different age groups. The results suggested that strong guidance had a positive impact on students' development of metacognitive thinking skills independently of their age and gender.

## Group metacognition

There is a recent shift in the literature towards the study of group metacognition rather than metacognition as an individual endeavor. Yet, research on metacognition in group (e.g., CSCL) situations is not well developed, despite group learning being commonplace in schools and other learning environments (Smith & Mancy, 2018). Despite the limited research on group metacognition, some findings suggest that metacognition is mediated and socially shared among group members in collaborative activities (Goos, Gailbraith & Renshaw, 2002) and that group metacognition can be considered as an extension of individual metacognition into group interactions. Also, researchers agree that metacognition in group situations consist of students' monitoring, reflecting and controlling of one-another's knowledge and actions. In CSCL research, the potential role of CSCL tools for supporting group metacognition has not been examined to date. As Järvelä and Hadwin (2013) explained, the potential role of CSCL tools for supporting the planning, monitoring, and regulation of collaborative learning processes has been virtually ignored. In this work, we see ER as CSCL tools that can promote students' metacognitive thinking. Most of ER activities are collaborative learning activities, yet, there are virtually no studies in the CSCL literature that examine the impact of ER on the development of group metacognition as an essential part of group work.

## Methods

### Participants

The participants were 14 students (6 male and 8 female) in Grades 4, 5 and 6 (aged 9-11 years old) in a public primary school in Cyprus. The students worked in 4 groups of 3-4 students each. Each group was formed with different genders and abilities (i.e., mathematical, technological and problem-solving abilities) to allow different discourses and problem-solving approaches to develop. The participating students had no previous experience in robotics.

### Procedures

There were two weeks of introductory activities to help students get familiar with the EV3 kit. These activities were followed by three 80-minutes sessions of STEM-related problem-solving tasks. Students should program a robot using a tablet, which was connected to the robot via Bluetooth. Each group was tasked with the following programming problems.

- Program a robot to move from its starting position, through a maze, to the finish position
- Program a robot to move along the outside of the flags without touching them
- Program a robot to draw a hexagon

Students in groups could adopt any approach they wanted to come to a solution. The teacher acted as a facilitator assisting the whole procedure e.g., assessing progress, examining understandings, monitoring group



work, and suggesting attention to data. When the groups completed their tasks, a debriefing phase took place. In this case, the groups demonstrated their approach in addressing the problem and answered questions asked by the facilitator and the students of other groups.

### Data collection and analysis

Verbal contributions were recorded via audio recorders next to each group. A camera was also placed in the room to record the overall student interaction and technology use.

To answer RQ1, the audio data were transcribed verbatim and analyzed using a fine-grained analysis. The unit of analysis was the individual participant and the discourse was coded on a turn-by-turn basis. A new turn was considered to start when the speaker changed. When the speaker shifted the theme of the discussion or when a different kind of discourse appeared, these were parsed into extra coded units. Generally, a conversational turn had more than one coding units. For instance, when a student asked a question but also added one or more statements, this was coded as two or more different coding units. Two independent raters coded 35% of the data to verify the reliability of coding. Reliability was acceptable (agreement 75%), and therefore, the first researcher completed coding the complete dataset. We used the coding scheme reported in Hmelo-Silver (2003), which conceptualizes the thinking processes and the general cognitive, metacognitive and social characteristics involved in collaborative knowledge construction.

To answer RQ2 and RQ3, a group was selected for further examination with a chronological investigation of within-group interaction. We used the CORDTRA technique initially presented by Hmelo-Silver, Jordan, Liu, and Chernobilsky (2011) and later applied in varied CSCL settings by Ioannou (2011), Ioannou, Brown and Artino (2015), and Socratous and Ioannou (2018). CORDTRA was examined in combination with excerpts of students' discourse to identify details of metacognitive and collaborative processes and the role of the technology.

## Findings

### What are the elements of collaborative knowledge construction? (RQ1)

#### Knowledge

As presented in Table 1, the students rarely referred to prior conceptual knowledge or experience of knowledge (5.2%). Not surprisingly, the students made comparisons and links referring to observations of previous actions in the same task.

#### Metacognition

The students used a larger amount of metacognitive utterances (24.1%). The majority were monitoring statements. Planning contributions occupied the second largest percentage of metacognitive utterances and were almost always in response to data derived from the results of previous trials. However, students did not mediate their planning with, prior knowledge, experience, or existing theories.

#### Interpretation

Students dedicated some effort in interpreting data derived from the robot or the tablet display (6.8%). Interpreting data was an opportunity to reconsider, test and refine their solutions.

#### Collaboration

Collaboration category included three subcategories: conflict, questioning and facilitator's input. Conflicts within groups were few and appeared mostly at early stages of the task. When conflicts appeared, they were more often related to the robot's failure to perform the expected outcomes; conflicts were rarely over a concept. Student groups generated many questions (22.7%), most of which referred to teammates rather than the facilitator. Most of these questions were planning-related questions as well as software- and robot-related questions. As shown by the relatively large number of statements related to agreement with peers (12.8%), students' consensus-seeking behavior was frequent. Responses by the students (24.3%) revealed the degree of consensus within the group. Students constructed simple explanations and brief answers more often than they elaborated explanations. Facilitator questioning was mainly concerned with software and robot use. The main operation of the facilitator's input was coded as monitoring (7.4%).

Table 1: Major categories and subcategories frequencies

| Coding categories    | N (%)            | Coding categories    | N (%)              |
|----------------------|------------------|----------------------|--------------------|
| <b>Knowledge</b>     | <b>29 (5.2%)</b> | <b>Collaboration</b> | <b>355 (63.9%)</b> |
| Conceptual knowledge | 5 (0.9%)         | Conflict             | 26 (4.8%)          |

|                        |                    |                            |             |
|------------------------|--------------------|----------------------------|-------------|
| Prior experiences      | 4 (0.7%)           | Conceptual                 | 3 (0.6%)    |
| Analogies              | 20 (3.6%)          | Task-specific              | 23 (4.1%)   |
| <b>Metacognition</b>   | <b>134 (24.1%)</b> | Questioning                | 126 (22.7%) |
| Monitoring             | 74 (13.3%)         | Clarifications             | 30 (5.4%)   |
| Evaluation             | 4 (0.7%)           | Plan-related               | 44 (7.9%)   |
| Reflection             | 19 (3.4%)          | Software-related           | 22 (4%)     |
| Total planning         | 37 (6.7%)          | Self-answered              | 5 (0.9%)    |
| Theory-driven Planning | 2 (0.4%)           | General                    | 2 (0.4%)    |
| Data-driven Planning   | 33 (5.9%)          | Facilitator                | 23 (4.1%)   |
| Unjustified            | 2 (0.4%)           | Responses                  | 135 (24.3%) |
| <b>Interpretation</b>  | <b>38 (6.8%)</b>   | Agreement with facilitator | 20 (3.6%)   |
| High-level             | 7 (1.3%)           | Agreement with peer        | 71 (12.8)   |
| Low-level              | 31 (5.5%)          | Brief answers              | 24 (4.3%)   |
|                        |                    | Simple explanations        | 16 (2.9%)   |
|                        |                    | Elaborate explanations     | 4 (0.7%)    |
|                        |                    | Facilitator's input        | 68 (12.1%)  |
|                        |                    | Monitoring                 | 41 (7.4%)   |
|                        |                    | Explaining concepts        | 3 (0.4%)    |
|                        |                    | Explaining Software        | 24 (4.3%)   |

### How does ER help to activate group metacognitive processes? (RQ2)

To answer RQ2, we examined an integrated view of an episode, using the CORDTRA diagram of Fig. 1. On CORDTRA diagram the numbers on the x-axis represent the chronological order of the coded units, whilst the y-axis represents the coded categories (records 6 to 29) and the speakers (records 1 to 5). The diagram reveals the nature of student's talk, including metacognitive talk, and its temporal relation to the use of the robot when students tried to solve the "Draw a hexagon" challenge. Combining the diagram with discourse excerpts helped to understand student's interactions across time. We zoomed into an episode in which the students work on solving the "draw a hexagon" challenge (lines 90-190).

#### ER activating group metacognitive processes through embodied interaction

For the activity, students should combine mathematical knowledge, experience with ER (i.e., introductory lessons), and programming skills to solve the problem. First, students started to discuss how they could solve the problem without having many ideas. A student stated that they should use the gyro sensor while another student added that they should place a pen holder on the robot. A detailed discussion about where they could set the pen holder took place in lines 91-99. Here, questioning discourse appeared as an essential aspect of collaborative knowledge construction. The students' questioning about where to put the gyro sensor and the kinds of turns the robot should make, triggered the dialogue for the next steps. The students started to research the question using the robot as a mean for experimentation by adjusting the pen holder in different places on the robot. Students seemed to recognize the significance of where they should adjust the pen holder; this important discussion moved students' thinking forward. The overall experimentation involved their bodies as students held the robot in their hands and were trying to simulate (with their bodies) possible movements of the robot and thinking of possible pen footprints on the paper. Students involved their bodies in understanding the difference between swing and point turn (lines 100-114). Students' embodied interaction with the physical robot triggered further social interaction and stimulated group metacognitive processes. The students tested and modified their new ideas, against existing knowledge and new data. Thus, it appears that ER, through embodied interaction, served as a tool for experimentation, activating group metacognitive processes and collaborative knowledge construction.

- Student 1 We must draw a hexagon (*laughing*). Any ideas?
- Student 3 We must use the gyro sensor to turn exactly as degrees as we program it (*for accurate angle measure*).
- Student 2 Yes, the robot must turn exactly as degrees as we program it.
- Student 4 We also need to adjust a marker to draw the hexagon as the robot moves and turns.
- Student 2 What kind of turns?
- Student 3 Turns.
- Student 1 Pivot turns. The robot must turn very sharp and make pivot turns to draw an angle.

Student 1 Ok! Where can we apply the marker?

Student 2 If we put it here (*holds marker and robot and try to make turns to draw a random angle*).

*These tests continued until students manage to draw angles formed by two rays rather than curved lines. Students placed the marker on different parts of the robot and tried to draw random angles. They tried to put the marker in different places (between the wheels, next to the right wheel and on the back of the robot) to understand where it would be more efficient to place the marker.*

### ER activating students' group metacognitive processes through interactivity and transparency

In lines 160-176 of the CORDTRA diagram, students went through an exploration in which they used their conceptual knowledge of mathematics and programming in a real-world situation. Students were concerned about how many degrees their robot should turn and, with the teacher's assistance, they managed to connect their mathematical knowledge and programming skills with a real-world condition. A student influenced by the introductory robotics lessons used a flowchart describing the required moves of the robot to draw a hexagon (line 168). Then, they decided to program the robot to turn  $120^\circ$ , as much as the internal angle and observed their robot turning much more than they expected. Immediate feedback from the robot's moves (i.e., observing the robot turn more than they expected) made the students think and monitor their thoughts (line 169). Robot's failure to produce the expected outcome seems to have triggered the group's metacognitive thinking. Thinking of what they were doing wrong, checking various aspects (lines 169-172) and building on each-other's thoughts, they excluded various possibilities and proposed a solution to the problem. After that, student 4 contributes a more advanced thinking to the discussion, suggesting that they should put a smaller value for the turning angle because with  $120^\circ$  the robot was turning too much. Student 4 proposed to represent the problem on a paper to calculate the turning angle. Students acknowledged this idea and began to model the problem on a paper. By representing the problem on paper students managed to find the correct value for the turning angle. Then, Student 1 made his thinking visible showing on the paper the correct angle (line 174). Student 2 built on the previous thought proposing the solution to the problem (line 175). Therefore, the process of socially-shared metacognition emerged in this group when Student 4 provided a metacognitive regulation statement (i.e., "If the robot turns  $120^\circ$  left, it will get into the hexagon. Let's draw the hexagon on a paper to find the angle"). The other group members acknowledged this contribution and developed a solution to the problem.

The transparency features of ER helped the students think, apply and check their ideas to overcome the problem. Easy changes to the software and hardware, at no cost, helped the students to avoid frustration, and through the open and accessible programmability of the robot, they managed to overcome the obstacles. The robot's programming, the expected results, and the actual results of its actions served as a metacognitive tool and as a data reference that students could use to negotiate their developing solution. The students identified gaps in their knowledge and collectively discussed, elaborated, and improved their solution. Regulatory statements that were produced due to the interactivity and transparency features of the robot promoted group metacognition and facilitated collaborative knowledge construction.

Student 3 Now, we will program the robot to move forward, then make a turn for some degrees then again forward and then turn, etc.

Student 2 Ok, we have to think about how many turns and how many degrees.

Student 1 Six turns and six forward. I do not know how many degrees.

Teacher What do we know about the total internal angles of polygons?

Student 4 It depends on how many different triangles are formed into the hexagon that does not overlap each other.

Student 2 How many different triangles does a hexagon have?

Student 3 I will draw a hexagon to find how many triangles are formed.

Student 3 4 different triangles. So, multiplies by  $180^\circ$  each equal  $720^\circ$

Student 1 Divide by 6 angles of a hexagon (*thinking*). Equals  $120^\circ$ . So, we will program the robot to move forward and then turn  $120^\circ$  for 6 times.

*(The team programmed the robot and is going to test the program).*

Student 2 No, it is turning too much. Perhaps we calculate the angles wrongly. Let's check it.

Student 3 Or, the sensor is not working

Student 2 Gyro sensor looks ok!

- Student 1 *(They are doing the calculations)* The angle is correct 120°. Must be something else.
- Student 4 Yeah, but I think we just have to take a smaller angle. 120° are all the internal angles of the hexagon. The robot moves on one of the sides of the hexagon. If the robot turn 120° left, it will get into the hexagon. Let's draw the hexagon on a paper to find the angle.
- (They draw a hexagon with a robot, representing it with a dot, on one of its angles)*
- Student 1 The robot is this dot and must turn here *(showing with his finger)*. So the turning angle is this one, we must find this one *(showing on the paper)*.
- Student 2 This angle is supplementary of the internal angle. So its 180-120 = 60.
- Student 4 Yes, that is. The robot must turn as much as the supplementary of the internal angle, only 60° not 120°.

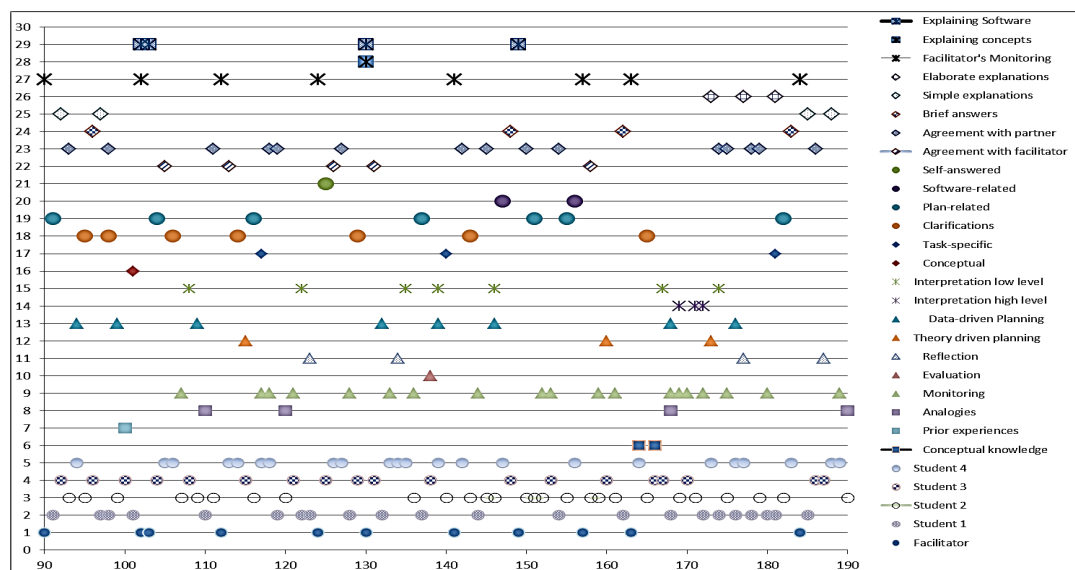


Figure 1. CORDTRA diagram of students' contributions.

### What is the relationship between collaborative talk and metacognitive talk? (RQ3)

Both in the previous excerpt and the one below, the students made their metacognitive thinking visible mainly in mutual interaction with their teammates. Student 1 tried to explain their failure to solve the challenge proposing that the flags were small, so the sensor could not detect them. This contribution triggered the thinking of Student 3, leading him to suggest a new idea that is, the use of two ultrasonic sensors instead of one. Student 1 pointed out his disagreement over the proposed idea and documented his position using the experience of a previous failing effort outside the current activity. Then, Student 1 contributed a metacognitive statement to justify his position proposing that they do not know well how to handle an ultrasonic sensor and so, he proposed a trial and error plan. Student 3 ignored Student's 1 plan highlighting that with two sensors, it would be easier for the robot to detect the flags. When they failed, Student 3 accepted to use the alternative plan but he first proposed to measure the distances among the flags so that they did not use a trial and error plan. In the excerpt below, the students compared their thinking with the thinking of their peers and this involved the use of collaborative talk in parallel with metacognitive talk. Also, as shown in the CORDTRA (Fig. 1), collaborative and metacognitive talk seemed to have a temporal relationship between them. For example, contributions that were coded as collaborative talk were usually followed by one or more metacognitive contributions vice versa.

- Student 2 We will use the ultrasonic sensor to avoid the flags.
- Student 3 Ok then. Put the ultrasonic sensor. *(They executed their plan, but they failed)*.
- Student 1 The flags are small, so the sensor cannot detect them.
- Student 3 We can use two ultrasonic sensors. What do you think?
- Student 1 No, we tried to use the ultrasonic sensor once, and we failed. We do not know how to handle it. Let's program the robot to move, and then we can adjust the values.
- Student 3 If we put two sensors, it will be easier for the robot to detect the flags.

- Student 2    Ok! Let's try with two sensors.
- Student 1    Ok then. (*They executed their plan using two sensors, but they failed*).
- Student 1    I told you, we do not need the sensors.
- Student 3    One more trial with two sensors and then, if we fail, we can move with your plan.  
(*They changed the position of the two sensors and tried again, but they failed*).
- Student 3    Ok. Let's do what you said, but first, we can measure the distance between the flags to calculate the value of rotations.

## Discussion

The study presents evidence that CSCL activities using ER can engage students in collaborative knowledge construction with prevalent elements of metacognitive processes, questioning, and answering. Indeed, students' discourse demonstrated logical reasoning coupled with metacognitive statements enabling the students to predict and to plan the flow of actions required to solve the problem. Monitoring elements of metacognition seem to be activated in an ER learning environment, engaging students in the process of exploration for the acquisition of knowledge. The large volume of monitoring elements of metacognition can be explained as the ER's value in encouraging procedural knowledge rather than declarative knowledge i.e., student learning by doing and understanding strategies of problem-solving rather than concepts.

During the ER activity, intensive collaboration was enacted in the form of questioning and answering while metacognition was enacted in the form of monitoring and planning. Many researchers have identified questioning (e.g., Hmelo-Silver & Barrows, 2008) and reflective thinking (e.g., Baker & Lund, 1997) as important kinds of discourse in knowledge building situations. Contributions of prior knowledge were limited, although this might not be replicated in a setting where learners have prior experiences with ER. Our findings confirm previous evidence about ER promoting collaborative knowledge construction (Chambers, Carbonaro, Rex & Grove, 2007; Socratous & Ioannou, 2018). This work contributes further in that it presents a fine-grained analysis of the phenomenon to strengthen the scientific evidence in the area. While previous studies rely heavily on the study of metacognition as an individual endeavor, using self-reported data (e.g., Atmatzidou et al., 2018), this study documents metacognition as a result of group work, while it occurred in-situ.

Metacognitive elements, coded as monitoring, evaluation, reflection, and planning, are activated in ER activities through embodied interaction with the physical robot. Indeed, when a robot is being used in the activity, it enables students' physical action and simulation of the robot's expected actions. Such activities seem to encourage expression and personal involvement in the learning process, whilst supporting teamwork which is important for the metacognitive process. Moreover, the transparent software design and the direct interactivity (feedback) coming from the robot's moves in response to students' programming, seem to facilitate the group's metacognitive thinking. In fact, when the robot failed to perform the expected outcomes, monitoring and planning elements of metacognition were documented on our chronological diagrams. Metacognition was necessary for students to understand how the tasks were performed and to be able to identify problems, negotiate modifications and operating changes to solve the problems. Embodied interaction with the physical robot, combined with feedback coming from the robot, acted as an extension of students' mind, scaffolding knowledge construction by re-evaluating their solutions. From this perspective, ER can be considered as "scaffolding embedded technological tools" (Chambers et al., 2007).

Our research has provided some initial evidence for a temporal relationship between collaborative talk and metacognitive talk in a problem-solving ER environment. The study further presents an instrumental role of ER technology in supporting metacognitive processes in CSCL settings. Metacognitive and collaborative talk appear to mediate each-other in this CSCL, ER setting. We understand that this evidence is not clear yet. Further development of our understanding of ER as metacognitive tools, will help us develop strategies to fully maximize their effectiveness in group problem-solving CSCL tasks.

## Conclusions

Coding and plotting student's discourse around an ER experience in CSCL settings can shed light on the value to the technology for collaborative knowledge construction and group metacognition. In this work, fine-grained analysis of student's discourse made evident that metacognition, along with questioning and answering, are prevalent elements of collaborative knowledge construction in ER activities. What is more, the role of the technology seems to be instrumental; namely, the embodied interaction, direct feedback and openly accessible programmability enabled by the robot were tightly coupled with group metacognitive processing and overall collaborative knowledge construction. In conclusion, this work extends the evidence on the value of ER integration in learning environments and CSCL activities. The study contributes in that it presents a fine-grained

analysis of the phenomenon to strengthen the scientific evidence in the area. While previous studies rely heavily on the study of metacognition as an individual endeavor using self-reported data (e.g., Atmatzidou et al., 2018), this study documents metacognition as a result of group work, while it occurred in-situ. Future work should extend on the nature of the problem, the teacher's scaffolding, the students' roles and the characteristics of the technology which might further endorse collaborative knowledge construction and metacognition in CSCL settings.

## References

- Ardito, G., Mosley, P., & Scollins, L. (2014). We, robot: Using robotics to promote collaborative and mathematics learning in a middle school classroom. *Middle Grades Research Journal*, 9(3), 73.
- Atmatzidou, S., Demetriadis, S., & Nika, P. (2018). How Does the Degree of Guidance Support Students' Metacognitive and Problem Solving Skills in Educational Robotics? *Journal of Science Education and Technology*, 27(1), 70-85.
- Baker, M., & Lund, K. (1997). Promoting reflective interactions in a CSCL environment. *Journal of computer assisted learning*, 13(3), 175-193.
- Bers, M. U., Flannery, L., Kazakoff, E. R., & Sullivan, A. (2014). Computational thinking and tinkering: Exploration of an early childhood robotics curriculum. *Computers & Education*, 72, 145-157.
- Chambers, J. M., Carbonaro, M., Rex, M., & Grove, S. (2007). Scaffolding knowledge construction through robotic technology: A middle school case study. *Electronic Jnl for the Integration of Technology in Education*, 6, 55-70.
- Constantinou, V., & Ioannou A. (2018). Development of Computational Thinking Skills through Educational Robotics. In European Conference on Technology Enhanced Learning - ECTEL. Retrieved December 2018 from <http://ceur-ws.org/Vol-2193/paper9.pdf>
- Efklides, A. (2011). Interactions of metacognition with motivation and affect in self-regulated learning: The MASRL model. *Educational psychologist*, 46(1), 6-25.
- Gaudiello, I., & Zibetti, E. (2013). Using control heuristics as a means to explore the educational potential of robotics kits. *Themes in Science and Technology Education*, 6(1), 15-28.
- Goos, M., Galbraith, P., & Renshaw, P. (2002). Socially mediated metacognition: Creating collaborative zones of proximal development in small group problem solving. *Educational studies in Mathematics*, 49(2), 193-223.
- Harel, I., & Papert, S. (Eds.). (1991). *Constructionism*. Westport, CT, US: Ablex Publishing.
- Hmelo-Silver, C. E. (2003). Analyzing collaborative knowledge construction: Multiple methods for integrated understanding. *Computers & Education*, 41(4), 397-420.
- Hmelo-Silver, C. E., & Barrows, H. S. (2008). Facilitating collaborative knowledge building. *Cognition and instruction*, 26(1), 48-94.
- Hmelo-Silver, C. E., Jordan, R., Liu, L., & Chernobilsky, E. (2011). Representational tools for understanding complex computer-supported collaborative learning environments. In *Analyzing interactions in CSCL* (pp. 83-106). Springer US.
- Ioannou, A. (2011). Online collaborative learning: The promise of wikis. *International Journal of Instructional Media*, 38(3), 213-223.
- Ioannou, A., Brown, S.W., & Artino, A.R. (2015). Wikis and forums for collaborative problem-based activity: A systematic comparison of learners' interactions. *The Internet and Higher Education*, 24, 35-45.
- Järvelä, S., & Hadwin, A. F. (2013). New frontiers: Regulating learning in CSCL. *Educational Psychologist*, 48(1), 25-39.
- Jonassen, D. H. (2000). *Computers as mindtools for schools: Engaging critical thinking*. Prentice Hall.
- Keren, G., & Fridin, M. (2014). Kindergarten Social Assistive Robot (KindSAR) for children's geometric thinking and metacognitive development in preschool education: A pilot study. *Computers in Human Behavior*, 35, 400-412.
- La Paglia, F., Caci, B., La Barbera, D., & Cardaci, M. (2010). Using robotics construction kits as metacognitive tools: A research in an Italian primary school. *Studies in Health Technology and Informatics*, 154, 110-114.
- Pintrich, P. R., Wolters, C. A., & Baxter, G. P. (2000). 2. assessing metacognition and self-regulated learning.
- Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational psychology review*, 7(4), 351-371.
- Smith, J. M., & Mancy, R. (2018). Exploring the relationship between metacognitive and collaborative talk during group mathematical problem-solving—what do we mean by collaborative metacognition?. *Research in Mathematics Education*, 20(1), 14-36.
- Socratous, C., & Ioannou, A. (2018). A Study of Collaborative Knowledge Construction in STEM via Educational Robotics. In J. Kay & R. Luckin (Eds.), *Rethinking Learning in the Digital Age: Making the Learning Sciences Count, 13th International Conference of the Learning Sciences (ICLS) 2018* (Vol. 1, pp. 496-503). London, UK: ISLS.

## Acknowledgments

This work is part of a project that has received funding from the European' s Horizon 2020 research and innovation programme under grant agreement N°739578 and the government of the Republic of Cyprus through the Directorate General for European Programmes, Coordination and Development.

# Disciplinary Task Models for Designing Classroom Orchestration: The Case of Data Visualization for Historical Inquiry

Josh Radinsky, University of Illinois Chicago, joshuar@uic.edu  
Iris Tabak, Ben-Gurion University of the Negev, itabak@bgu.ac.il  
Monika Moore, University of Illinois Chicago, mmoore49@uic.edu

**Abstract:** Incorporating data visualizations into college classrooms can enhance disciplinary learning, when accompanied by associated meaning-making practices. However, tools not specifically designed for CSCL may lack supports for collaboration and disciplinary reasoning. Therefore, it is necessary to design supports into classroom orchestration. In this paper we examine the design of such orchestration to scaffold historical reasoning with geospatial data visualizations. The orchestration of a sequence of small-group tasks scaffolded progressively more complex reasoning about historical waves of immigration, which the instructor leveraged to support students' reasoning at multiple levels in whole-class discussion. Our analyses point to these kinds of orchestrated scaffolds as a valuable direction for CSCL research and design.

In proposing productive research directions for CSCL (Ludvigsen, Cress, Law, Rosé, & Stahl, 2016; Wise & Schwarz, 2017), a number of scholars converge on broadening the purview to tools beyond the products of CSCL design research. Specifically, Reimann (in Ludvigsen et al., 2016), foresees greater attention to sophisticated representational systems used for everyday and professional goals. Ludvigsen (2016) further calls for research to identify practices for arriving at shared meanings through interactions with these representational systems, and studies that highlight when and why collaboration might be useful (Wise & Schwarz, 2017).

In this paper we explore ways to incorporate such systems and their accompanying knowledge-building practices into the post-secondary classroom. College classrooms are environments in which learners can be introduced to disciplinary inquiry for the first time. This includes being exposed to a discipline's *professional vision* (Goodwin, 1994) – being apprenticed by an instructor to see information (texts, data) through a particular lens that enables discipline-relevant insights and reasoning. This is a process that often proves challenging for college instructors (PCAST, 2012). Merely using representational tools in instruction does not guarantee that learners will see data through the eyes of the profession, especially when the tools lack support for collaboration and disciplinary meaning making (e.g., Fischer et al., 2013; Raes, Schellens, De Wever, & Vanderhoven, 2012). Incorporating these tools poses a number of design challenges: how to create coherence between existing curricula and the forms of thought introduced by the tools; how to help learners see and interpret representations through disciplinary eyes; and how to manage the complexity inherent to these tools. Meeting these challenges requires design work *around* these tools themselves.

We propose that integrating disciplinary models with *classroom orchestration* is a productive way to incorporate existing representational systems into instructional settings. We examine this in a college history education class using a data visualization tool. The study addresses the questions: How can data visualization tools be incorporated into the college classroom to support historical inquiry? How can classroom orchestration facilitate learners' ability to engage in geo-spatial reasoning for historical inquiry?

## Theoretical framework

### Classroom orchestration and distributed scaffolding

Research in CSCL and the Learning Sciences often argues that the meanings, purposes, and learning value of tools are mediated by the activity structures in which those tools are used. *Classroom orchestration* (Dillenbourg, Prieto, & Olsen, 2018) is a process by which teachers or other instructional designers plan, deploy and regulate multiple resources in the classroom to facilitate learning and achieve a productive workflow (Prieto, Holenko Dlab, Gutiérrez, Abdulwahed, & Balid, 2011). A central role for teachers in classroom orchestration is to assess progress and achievement, and make changes and adaptations on the fly to better align interactions with learning goals (Kaendler, Wiedmann, Rummel, & Spada, 2015). This alignment of multiple resources and interactions to support learning has been described as *distributed scaffolding* (Puntambekar & Kolodner, 2005; Tabak, 2004).

Classroom orchestration and distributed scaffolding can support learners in specific reasoning processes, such as learning to interpret data through a particular disciplinary lens, or *professional vision* (Goodwin, 1994). Members of a professional community are likely to converge on similar interpretations, while members of different communities may focus on different aspects and arrive at different meanings, even when observing the

same representation (Bowen, Roth, & McGinn, 1999). Learners may be able to approach novel representational systems and derive meaning from them, but they may struggle to produce the types of explanations that typify a particular field or discipline. Therefore, a pivotal role for distributed scaffolding and orchestration is in helping learners to focus on particular attributes, and to interpret representations in particular ways, such as by teachers interpreting representations in disciplinary terms alongside learners (Tabak, 2004; Tabak & Reiser, 2008).

### Moderating discussion through levels of visualization

Classroom orchestration can also involve regulating classroom discourse. Discussions around modeling or visualization tools can occur at three levels (Radinsky, Milz, Zellner, Pudlock, Witek, Hoch & Lyons, 2017). At the *interface level*, discussion centers on how to manipulate the tool and extract information and visualizations. At the *modeling level*, discussion centers on the signs, symbols, images and quantities that are represented in the tool and their relationship to the phenomena being represented. Discussions at this level can be about relationships among variables and the interpretation of patterns. At the *represented-world level*, discussion centers on the phenomena in the world that are modeled through the visualization. Discussions at the represented-world level arise from the representations, but speak in the language of the modeled phenomena rather than the images, symbols or quantities that are displayed. The goal is to push discussion beyond the interface level, to concentrate on the modeling level to ensure that new understandings are constructed, and then to the represented-world level so that the knowledge that is constructed is about the phenomena of interest. Although the instructional aim is to move from the language of representation to the language of historical phenomena, it is critical that classroom guidance regulate the connection between discussion at both levels, so that the reasoning stays anchored in data.

### Data visualizations and historical reasoning

Historical reasoning involves the construction of narratives of the past that are grounded in observations from historical records (Seixas, 2017). This requires skills such as chronological reasoning, contextualizing historical data, corroborating observations across multiple sources of information, and reasoning about texts as sources (Seixas, 2017). Using a data visualization tool for historical inquiry involves more than extracting and interpreting information; it requires viewing the data as historical records of people whose experiences can be only partially and imperfectly inferred from observations of these data. Geographic information systems (GIS) are a class of visualization tools that has been part of a fundamental spatial shift in many of the social and natural sciences, but has held a more problematic place in the discipline of history. Though some historians find an ontological or epistemological mismatch between the quantified and spatialized representations of GIS and the tentative and contingent nature of historians’ professional vision (Owens, 2007), there is a small but growing group of historians who embrace GIS tools because of their affordances for observing phenomena at spatial and temporal scales beyond what many traditional historical sources allow (Knowles, 2008). The present study explores a design for apprenticing students in historical reasoning that takes advantage of commonly-available GIS tools.

### Using disciplinary task models to design orchestration with a “webmap”

The visualization tool used in this study, *Immigration Explorer*, is what Baker (2005) describes as “webmaps”: online, public-use GIS providing access to geospatial historical data with minimal functionality for manipulating the representation, instead emphasizing ease of access and simplicity of use. This was a free resource available on a public-access URL created in 2009 by the *New York Times* for browsing historical census data for the years 1880-2000. The webmap displayed the total population of each county in each census decade; the number of people who were identified as foreign-born; and the number of foreign-born people from each of 23 countries-of-origin. The data were visualized as spheres of varying size, representing the number of county residents born in each country-of-origin. Figure 1 shows an example map of the census count of people born in Mexico living in each U. S. county in 1970, 1980 and 1990.

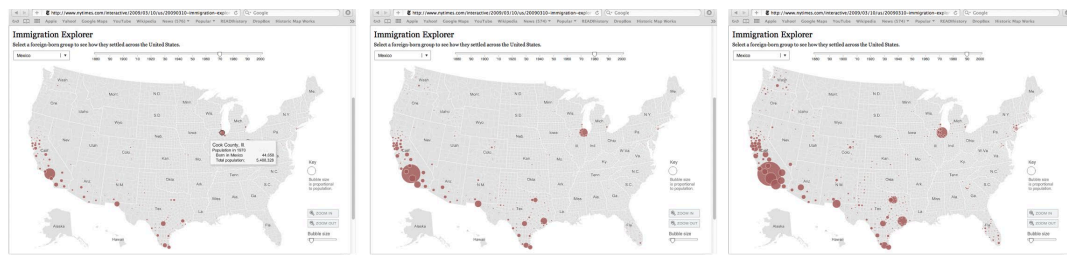


Figure 1. *Immigration Explorer* maps showing Mexican-born population by U. S. county, 1970-1990.



## The design: Orchestration of sequential collaborative data interpretation activities

The simplicity and accessibility of this tool allowed for ease of integration into the classroom but required any distributed scaffolding for historical reasoning to be designed into the orchestration of the classroom activity, since no such supports were built into the tool. The design of these “wrap-around” classroom supports, to foster reasoning that is both historical and geo-spatial, is the focus of the present analysis.

Learners required considerable support in order to engage in meaningful historical inquiry while developing nascent models of multiple population patterns shifting over time. Therefore, the tacit conceptual moves involved in this reasoning process were mapped onto group activity structures to design classroom orchestration. This provided learners with a series of manageable tasks that combined to form richer, more complex interpretations of historical data. The final products of the group work could subsequently be used by the instructor in a whole class discussion to piece together, through a series of prompts, queries, and restatements, the initial building blocks of an historical account. This laid the conceptual groundwork for their subsequent interpretation of a series of historical case studies of migration (beyond the scope of the present analysis).

Five phases of activity with different group configurations guided students toward using the webmap to construct historical narratives about patterns of migration from different countries-of-origin (hereafter, CoO):

1. **Whole-class introduction of the mapping tool:** whole-class discussion showing how to use *Immigration Explorer* (interface level), introducing historical census data as a resource for historical reasoning (modeling and represented-world levels).
2. **Small-group investigation of one CoO in one region:** in groups of three, students select one CoO group and make observations of that group’s changing population pattern in one region over time. For example, they might examine the changing Chinese-born population in northern California from 1880 to 1930.
3. **Same-group comparing changing patterns for the same CoO in different regions:** in the same groups of three, students compare the patterns of population change of the same CoO in one region with another region, e.g., comparing Chinese-born population patterns in northern California to the very different patterns of Chinese immigration in the Midwest, around Chicago.
4. **New “jigsaw” groups compare “waves of migration” for different CoO:** students form new groups of three that each include an “expert” on a different CoO population. In these “jigsaw” groups (Johnson & Johnson, 1982), students take turns teaching the patterns they observed for their population, and then consider similarities and differences across these CoO’s patterns of migration.
5. **Whole-class sharing and discussion of group findings:** groups present their findings and with the instructor’s help articulate connections and interpretations towards constructing historical narratives. Instructor scaffolds bridge from modeling level to represented-world reasoning.

At each of the small-group phases (see Figure 2) there are challenges for reasoning at all three levels (interface, modeling, and represented-world). Phase 2 allows a group to focus on only one region and one CoO as they practice with the tools, attempting to describe change over time. Phase 3 adds the challenge of doing this for two regions of the USA, but with the same CoO map. Phase 4 requires the use of three different CoO maps, and pushes them to reason about different spatio-temporal patterns for different immigrant groups, potentially over centuries.

Whole-class discussion led by the instructor (Phase 1) provides modeling of the language of careful and specific observations (e.g., “the population of people born in Mexico grew quickly in this area from 1970 to 1990”), prior to students practicing these kinds of observations in small groups (Phase 2), and prior to repeating this same skill while attending to multiple regions of the map (Phase 3). The decade slider provides support for noticing population changes without having to change maps, and both regions they are comparing are visible in the same map. Students then take responsibility for sharing those observations with a new group, and making similar comparisons across different maps displaying different CoO’s (Phase 4). Finally, the instructor-mediated whole-class discussion (Phase 5) allows for a more formal sharing of the cross-group observations and comparisons, allowing communal scrutiny of their findings. This affords scaffolding, modeling, and shaping of students’ language by the instructor, referring back to the worked example from Phase 1.

Though fairly straightforward as a “jigsaw” lesson, the rationale for this sequence embodies a set of propositions about learning that have value for CSCL researchers. As Vogel, Wecker, Kollar and Fischer (2017) note, “CSCL scripts are particularly effective for domain-specific learning when they prompt transactive activities (i.e., activities in which a learner’s reasoning builds on the contribution of a learning partner) and when they are combined with additional content-specific scaffolding (worked examples, concept maps, etc.)” (p. 477). The phases of activity reflect the conceptual steps that a historian might take in constructing an inquiry comparing waves of migration for different populations. For a single learner, conducting these steps would be taxing. By reducing the complexity of the task within the original group (comparing patterns within a single map display)

and then pooling resources by forming new comparison groupings, students are able to consider multiple, multi-level comparisons, without having to produce such complex comparisons from scratch. The whole-class, teacher-mediated sharing affords joint attention to reasoning across multiple regions, populations, and decades, modeling data not simply as numbers, but as representations of historical movements of people through space and time.

### Orchestrating Progressively More Complex Historical Observations

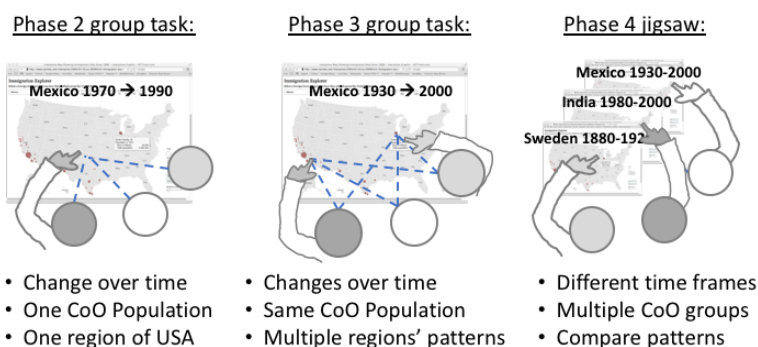


Figure 2. Orchestration “jigsaw” design constructs multiple, multi-level comparisons of historical migrations.

## Methods

### Participants and setting

Participants were 15 Masters students (12 female, 3 male) in a teaching licensure program. The lesson was conducted in a course taught by the first author, as an action research project, documented by a research assistant, and following IRB-approved procedures for consenting, data gathering and analysis. The course was a pre-service methods class for history and social studies teachers, and the focal lesson was part of a module on historical inquiry and migration. The larger instructional unit built conceptual understandings of migration, including specific cases of migration drawn from students’ families and a variety of texts, leading to an analysis of case studies from a historical text (not described here). The focus of this lesson was on different geographic and temporal patterns of immigration to the United States.

### Data collection and analysis

We report on a subset of the data collected for the larger research project: a video recording of the focal lesson (65 minutes). The classroom video was transcribed for speech and gesture by one research assistant, second-passed by another. All student names are pseudonyms. In addition, a pre-instruction reflection interview with the instructor (first author) was documented, articulating the logic of the sequence of group activities in the lesson. This was used to articulate the design rationale and intended trajectory of classroom talk.

The transcript of the final presentations and discussions was coded for descriptive and comparative discursive moves at the *interface*, *modeling*, and *represented-world* levels, following conventions described in (Radinsky et al, 2017), and a grounded-theory open-coding approach was used to identify the range of historical and geospatial descriptions, comparisons and explanations that emerged. This produced a set of 35 distinct codes, which are being developed into a coding scheme for ongoing analysis of project data. For the analysis presented here, qualitative descriptions of the unfolding spatial and historical observations, comparisons, and explanations were used to examine the ways the orchestration design supported instructional opportunities to scaffold disciplinary learning with the *Immigration Explorer* webmap.

## Findings

Due to space limitation, we present only details from the discussions that occurred in the whole-class segment following the jigsaw activity (Phase 5). We report on interactive presentations from two jigsaw groups: Lisa, Penny and Maritza, who studied immigrations from Vietnam, Sweden and Mexico, respectively; and Tina, Nancy and Cory, who studied immigrations from Sweden, Mexico and India, respectively.

### Lisa's group: Comparing quantities and rates (*modeling level*)

Lisa, Penny and Maritza presented each CoO group's pattern in order, starting with Vietnam. Lisa described how "we noticed the biggest county in 1980 was Orange County ... we looked at Orange County and Chicago for comparisons." Using these two counties in different regions of the country to construct their comparison, they noticed a pattern of the Vietnamese-born population doubling: "So from 1980 to 1990 it doubled in Orange County [California] and Cook County [Illinois, where Chicago is located]. And then from 1990 to 2000 it doubled also. Umm that was pretty much all from Vietnam."

When they moved on to present the data for Sweden, Penny followed Lisa's lead at the *modeling level* in presenting data for a single county (Douglas County, Minnesota), and then constructing a rate of change for that population in that county: "we went back to 1880 and we found um, a population of over 2,000 people ... And every 20, every 20 years ... every 20 years it decreased." Penny moves from naming quantities to giving a qualitative description of the changing map representation as she changes the time: "1940, we only have a little over 800, um, people in Douglas County and 20 years after by 1960 its in the 300s. And the bubble just continues to disappear and I don't find it, by the time we get to 2000, I don't find it. I don't find it anymore." Lisa then connected the patterns for the two groups: "So as you're – as the Swedish population in Minnesota was decreasing by half every 20 years the Vietnamese population was doubling every 10 years. So that was a cool comparison we thought of in our group."

Maritza's presentation of the pattern for people born in Mexico picks up on a different part of Lisa's initial description: the absence of data for some decades. She narrates their inquiry process as a series of discoveries of the limitations of the data, then shifts to add their observations about Hawaii: "we looked it up in Wikipedia that Hawaii didn't become a state until 1959. So that explains why there's no data here."

Three things are notable here: (1) their construction of the task of *making observations* as reporting single-county data values (rather than qualitative descriptions of regional patterns); (2) their use of data values to construct and compare rates of population change over time (doubling each decade, halving every 20 years); and (3) the emphasis on incomplete data. The instructor engages the exploration of some of these opportunities during the presentation, grounding the conversation in the *modeling level* (comparing quantitative patterns) that the group has highlighted. For example, when Maritza says "Once again, it [the data] disappears from Texas," the instructor adds: "And from the whole South. You see that? The whole South disappears," gesturing to the map to highlight the missing data Maritza has mentioned. This move takes up Maritza's language ("disappears"), adds to the observation in a way that better reflects the scale of the pattern ("the whole South"), and invites intersubjective attention to the data ("You see that?", gesturing).

A more active engagement occurs when Lisa presents the group's "cool comparison" of rates of change (doubling versus halving). The instructor endorses their positive self-evaluation ("Very cool"), and then asks the rest of the class to re-represent that comparison of rates-of-change:

- Instr: Very cool. Imagine of what ways we can visualize that, other ways we can show that change. Can anybody show that with your hands? What Lisa just said? Like if you were going to describe that change what would it look like?
- Penny: [points up with one hand and down with the other]
- Instr: So Penny went like this. [mimics Penny's hand motion, eliciting general laughter] What does this mean? [repeats gesture]
- Edita: Increase, decrease.
- Instr: Edita what are you going to show us again?
- Edita: [two flat hands, raising one and lowering the other] One decreased –
- Instr: Anyone got another one?
- Erica: Like this? [opens one hand and closes the other, eliciting general laughter]
- Instr: This is great. A way to visualize these numbers changing helps everybody.

This collaboratively-constructed sequence of re-representations of the group's reported data patterns (doubling versus halving) is explicitly endorsed as something that is "great" and "helps everybody." This presents an opportunity to reflect on multiple ways to communicate these phenomena, encouraging students to think beyond direct quantitative comparisons, in ways that support reflection on the phenomena they are modeling, and their own ability to model it in different ways.

## Tina's group: generating migration concepts (*represented-world* level)

Tina begins with a concise statement that bundles together a number of historical observations:

Tina: Our group we have Mexico, India and Sweden - Sweden. And then we said that - that Mexico has the most um, population compare to India and Sweden. And then in Sweden people actually stayed in the middle first but then um, India they come from California and New York, those port| areas first.

In contrast to the previous group, this statement foregoes a process of recounting each CoO's data separately and avoids reading out specific numbers for individual counties. Instead, it presents summary comparisons that include multiple populations: Mexico had more population than either India or Sweden; Swedish and Indian immigrants came to different geographic areas "first" ("the middle" for Sweden, versus "those ||port| areas" for India). This begins to bridge the modeling level with the represented-world level – using the patterns to begin to imagine the experiences of actual people in historical places. The instructor, sensing that there are multiple moves worth highlighting here, invites careful attention to Tina's observations:

Instr: I want everyone to get what Tina is saying. I want everyone to get this. Can someone repeat what she said? [no one volunteers immediately] [to Tina] OK, will you come up please and show us? I want everyone to get it. I don't want to let her off the hook until we all understand. She is comparing patterns now and she is not just using the numbers. So listen to how she describes it. So I've got Sweden right here. You want to do this? [motioning her to come up to the projecting computer]

Tina: [coming to computer] Alright, OK. So Sweden [selects Sweden, 1880] I think in the beginning is all in the middle [circles data pattern on display with mouse] but then if we compare to India [changes menu to India] there's no data over here until 1970 [changes year to 1970]. And then they all in the coastal areas - is that how you say it? Like New York [circling northeast data pattern] and California [circling southwest data pattern] compared to uh, Sweden again [selects Sweden, changes year back to 1880].

Tina: In the middle, in the middle [circling data pattern in central Midwest]

Instr: And you said "at the beginning"

Tina: Yes

Instr: So you compare when they *start* to show up, even though that's in different years.

Having brought Tina's group's complex spatial observations into the shared space of the projected display, the instructor again invites other students to re-represent it:

Instr: Can someone please restate what was the comparison that Tina made ... - the immigration patterns from people from Sweden, India and Mexico. How did she compare them? ...

Nate: So she compared them by uh, genesis point I guess. So when their populations first started to show up on the map, the sort of concentrations.

Instr: Can someone explain what Nate means by genesis point? He just gave us an awesome phrase that we can use. What is a genesis point for immigration? He's going to copyright it quickly so let's make sure we know. What does he mean by genesis point? Do you guys get this? If not please ask Nate to explain himself.

Beth: Like where they moved? Where they moved to? Entry point

Instr: Beth, say that again

Beth: Their entry point

Instr: Their entry point. Anyone else have another way of saying it?

Marie: The beginning of the census for uh, that specific region group um, on the map

Instr: The beginning of that census group on the map. You guys got it? Ok, so now ... so you said they each had a different genesis point.

Nate: Yeah, even though that was separated by years their uh, they had starting populations in areas and then spread out from there.

Several classmates are encouraged here to develop multiple ways of describing the phenomenon Tina's group introduced. The instructor's moves push the students beyond simply describing the patterns as midwestern and coastal, which locate more within the modeling context. By emphasizing her words "at the beginning," he brings out an important concept for historical reasoning about migrations: "So you compare when they *start* to show up, even though that's in different years." This moves the discussion from the modeling level to the represented-world level by pushing the comparison toward language that is relevant to phenomena of migration, leading to the development and uptake of Nate's concept of "genesis point" for an immigrant population, which is what enables Tina's group to make a geospatial comparison of the Swedish and Indian immigration patterns, despite their being separated in time by a century. Nate explains, "even though that was separated by years ... they had starting populations in areas and then spread out from there." As the class continues to examine this idea, Erica shifts to the represented world, connecting these patterns and her prior knowledge to a possible, partial explanation: "It seems like India's might be connected to hub airports because you have like Chicago, and is that Detroit?"

## Discussion

The learning opportunities created in this lesson proceed from the seeding of historical and spatial reasoning practices embedded in the progression of small-group activities. Like any CSCL design, the space of learning opportunities is mediated by the real-time decisions of the teacher, but the design of the learning environment provides distributed scaffolds and other resources that shape opportunities for disciplinary learning.

### Scaffolding multi-level disciplinary reasoning in group activity structures

All groups were able to bring back to the whole-class discussion descriptions of different CoO groups' changing population patterns. As the analysis demonstrates, there were notable differences between the two focal groups' representational practices in their presentations. Lisa's group remained mostly at the *modeling* level of reporting numbers and comparing quantitative changes for individual counties – valuable visualization practices that were taken up by the instructor and used to scaffold multiple representational moves in the classroom. Tina's group's comparisons of data patterns moved from the *modeling* level to the *represented-world*, enabling the instructor to engage the class in generating concepts for comparing regional migration patterns, even at different points in time.

Neither of these learning opportunities would have been likely in a lesson using this kind of webmap without the progression of distributed scaffolds orchestrated in the lesson. Although not all groups were able to provide descriptions that could be readily used to construct a historical narrative, enough groups were able to do so in order to provide the instructor with fodder to demonstrate how to use their observations in taking this next step. The instructor drew on the building blocks generated by the jigsaw groups to reposition their observations from discussions at the modeling level to discussions at the represented-world level. The progression from group presentations to the seeds of historical narratives in the culminating discussion suggests that the orchestration design, mapping multiple, multi-level comparisons into a sequence of group tasks, was effective in managing complexity and enabling novices to co-construct historical interpretations of complex visual data.

### Re-representation of patterns for disciplinary reasoning beyond the tool

Our analysis points to a promising strategy for fostering deep learning when using visualization tools. The aim of teaching with visualization tools is to move beyond the language of numbers, shapes or colors, to narrate the real-world phenomena they are meant to represent, so that learners develop knowledge about the phenomenon that endures even in the absence of the visualization tool. The interactive discussion here built on the jigsaw activity to encourage students to re-represent not only the data, but the patterns and phenomena they were co-constructing to compare migrations, both verbally and gesturally. Asking learners to re-represent patterns and concepts prompted co-construction of the ideas being discussed, free of the complexity of the data interface, while providing opportunities for formative assessment to guide the ongoing lesson. The outcomes of this strategy seen here suggest the value of further examining its potential for supporting deep learning and transfer.

## Conclusion

Following recent calls by leaders in CSCL (Ludvigsen et al., 2016; Wise & Schwarz, 2017) we examined how the design of classroom orchestration could reap the educational potential of sophisticated representational systems, which are used in professional and everyday contexts, in the college classroom. We propose that disciplinary task models can serve as a guide for making orchestration design decisions. We demonstrated how mapping configurations of group work to the conceptual reasoning moves that professionals make in using these tools could lead to disciplinary learning goals while using everyday tools like webmaps that were not designed as learning environments. Understanding how to orchestrate the classroom to cultivate students' disciplinary

thinking with tools that may not have built-in collaboration and reasoning scaffolds holds promise as a valuable area to explore in future CSCL work. Focusing on such tools may offer advantages beyond the support of disciplinary learning. Various visualization tools are becoming prevalent in daily life in online news, social networking, and other sites, making competency in interpreting these visualizations an important skill for civic participation. Using such tools in their everyday form in the classroom may lead to better transfer to everyday settings. In addition, as Roschelle et al have argued (2008), an alternative strategy for assuring the products of CSCL research have impact at scale is that rather than “scale up,” CSCL innovations be integrated into materials and tools that are already used at scale. Using disciplinary models to design orchestration around tools that are already widely available may be a way to achieve advances in college disciplinary learning at scale.

## References

- Bowen, G. M., Roth, W. M., & McGinn, M. K. (1999). Interpretations of graphs by university biology students and practicing scientists: Toward a social practice view of scientific representation practices. *Journal of Research in Science Teaching*, 36(9), 1020-1043.
- Dillenbourg, P., Prieto, L. P., & Olsen, J. K. (2018). Classroom orchestration. In *International handbook of the learning sciences*: Routledge.
- Fischer, F., Kollar, I., Weinberger, A., Stegmann, K., Wecker, C., & Zottmann, J. (2013). Collaboration scripts in CSCL. *The international handbook of collaborative learning*, 403-419.
- Goodwin, C. (1994). Professional vision. *American Anthropologist*, 96(3), 606-633.
- Kaendler, C., Wiedmann, M., Rummel, N., & Spada, H. (2015). Teacher competencies for the implementation of collaborative learning in the classroom. *Educational Psychology Review*, 27(3), 505-536.
- Knowles, A. K. (2008). Gis and history. In A. K. Knowles & A. Hillier (Eds.), *Placing history: How maps, spatial data, and gis are changing historical scholarship* (pp. 1-25). Redlands, CA: ESRI, Inc.
- Ludvigsen, S., Cress, U., Law, N., Rosé, C. P., & Stahl, G. (2016). Future-looking conversations in cscl. *International Journal of Computer-Supported Collaborative Learning*, 11(3), 255-262.
- Owens, J. B. (2007). Toward a geographically-integrated, connected world history: Employing geographic information systems (GIS). *History Compass*, 5(6), 2014-2040.
- President's Council of Advisors on Science & Technology Strategy to Improve K–12 STEM Education. (2012). Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics. Washington, DC: Executive Office of the President.
- Prieto, L. P., Holenko Dlab, M., Gutiérrez, I., Abdulwahed, M., & Balid, W. (2011). Orchestrating technology enhanced learning: A literature review and a conceptual framework. *International Journal of Technology Enhanced Learning*, 3(6), 583-598.
- Puntambekar, S., & Kolodner, J. L. (2005). Toward implementing distributed scaffolding: Helping students learn science from design. *Journal of Research in Science Teaching*, 42(2), 185-217. doi:10.1002/tea.20048
- Radinsky, J., Milz, D., Zellner, M., Pudlock, K., Witek, C., Hoch, C., & Lyons, L. (2017). How planners and stakeholders learn with visualization tools: using learning sciences methods to examine planning processes. *Journal of Environmental Planning and Management*, 60(7), 1296-1323.
- Raes, A., Schellens, T., De Wever, B., & Vanderhoven, E. (2012). Scaffolding information problem solving in web-based collaborative inquiry learning. *Computers & Education*, 59(1), 82-94.
- Roschelle, J., Tatar, D., & Kaput, J. (2008). Getting to scale with innovations that deeply restructure how students come to know mathematics. In E. K. Anthony, R. A. Lesh, & J. Y. Baek (Eds.), *Handbook of design research methods in education* (pp. 369-395). New York, NY: Routledge.
- Seixas, P. (2017). A model of historical thinking. *Educational Philosophy and Theory*, 49(6), 593-605.
- Tabak, I. (2004). Synergy: A complement to emerging patterns of distributed scaffolding. *Journal of the Learning Sciences*, 13(3), 305-335.
- Tabak, I., & Reiser, B. (2008). Software-realized inquiry support for cultivating a disciplinary stance. *Pragmatics & Cognition*, 16(2), 307-355.
- Vogel, F., Wecker, C., Kollar, I., & Fischer, F. (2017). Socio-cognitive scaffolding with computer-supported collaboration scripts: A meta-analysis. *Educational Psychology Review*, 29(3), 477-511.
- Wise, A. F., & Schwarz, B. B. (2017). Visions of CSCL: Eight provocations for the future of the field. *International Journal of Computer-Supported Collaborative Learning*, 12(4), 423-467.

## Acknowledgments

Funded in part by the I-CORE Program of the Planning and Budgeting Committee (LINKS center) and The Israel Science Foundation grant 1716/12, and by NSF grants DRL-0953448 and INSPIRE-1248052.

# Data Wrangling Practices and Process in Modeling Family Migration Narratives with Big Data Visualization Technologies

Shiyang Jiang, Carnegie Mellon University, shiyangj@cs.cmu.edu  
Jennifer B. Kahn, University of Miami, jkahnthorne@miami.edu

**Abstract:** Big data technologies are powerful tools for telling evidence-based narratives about oneself and the world. In this paper, we examine the sociotechnical practices of data wrangling—strategies for selecting and managing datasets to produce a model and story in a big data interface—for youth assembling models and stories about family migration using interactive data visualization tools. Through interaction analysis of video data, we identified ten data wrangling practices and developed a conceptual model of the data wrangling process that contains four interrelated recursive stages. These data wrangling practices and the process of data wrangling are important to understand for supporting future data science education opportunities that facilitate learning and discussion about scientific and socioeconomic issues. This study also sheds light on how the family migration modeling context positioned the youth as having agency and authority over big data.

**Keywords:** data wrangling, modeling, storytelling, family migration, data science education

## Introduction

As the interdisciplinary field of data science has grown, large-scale datasets, also known as big data, and interactive data visualization tools have become increasingly open and accessible, creating opportunities for learning. Assembling models and narratives with big data is an important STEM practice for many professions (Kosara & Mackinlay, 2013) and for participating in civic discourse (Philip, Schuler-Brown, & Way, 2013). The role of big data and visualizations in public conversations can be attributed to these technologies' capacities to support critical inquiry (boyd & Crawford, 2012). These tools are especially powerful for enriching narratives about the social and scientific world and providing evidence to challenge or confront misinformation.

Opportunities for youth to engage in data science activities that support analyses of social and scientific issues from local and global perspectives are needed. However, few studies have closely investigated how youth engage with big data interfaces at the level of discursive and embodied interaction. In this paper, we examine the sociotechnical practices of data wrangling (introduced in Kahn, under review), the strategies for selecting and managing datasets, for youth assembling models and stories about family migration with public big data using interactive visualization tools. We consider data wrangling as necessary for storytelling and modeling with big data, and an examination of the active interactions between youth modelers and visualization tools is necessary for understanding data wrangling. Our analysis addresses the following research questions: 1) *What are the **enactive practices** that describe participants' data wrangling interactions with big data visualization tools?* 2) *What is the **conceptual process** of data wrangling to assemble family data storylines?*

## Theoretical framework

First, this research follows an interactionist perspective (Greeno, 1994) that views learning as occurring through participation situated in activity (Bandura, 1986; Greeno & Engestrom, 2014; Lave & Wenger, 1991). Second, we approach family storytelling with big data as a representational, sociotechnical activity that involves coordination of bodies and tools (Hall & Nemirovsky, 2012). Third, learners' capacities, dispositions, and histories are central in our learning environment design and study of learning processes.

Furthermore, our study builds on a collection of data science education studies that have found that interactive, multivariable data visualization tools support learning across STEM and social studies disciplines (e.g., Philip, Olivares-Pasillas, & Rocha, 2016; Polman & Hope, 2014; Radinsky, Hospelhorn, Melendez, Riel, & Washington, 2014; Rubel, Hall-Wieckert, & Lim, 2017) and understandings of statistical concepts (e.g., Engel, 2017; Harrison, Yang, Franconeri, & Chan, 2014). These studies also call for more opportunities for youth to develop critical data literacy by using visualization tools to ask/answer questions and make inferences about data in personally and culturally meaningful ways (Borner, Peppler, Kennedy, Uzzo, & Heimlich, 2017).

## Methods

### Context and participants

We report on a single iteration of a design-based research program focused on exploring how youth and young adults learn to tell stories and build models about social and scientific issues with big data. Previous iterations found that making personal connections to large-scale phenomena represented by big data could be productive for generating critical perspectives. In turn, this design iteration sought to explore the benefit of embedding storytelling and modeling with big data in a personal context and to better understand the role of big data interfaces in relating personal experiences to a larger scale—the scale of the phenomena being represented by the big data. Conceptually, our design was thus intended to support learning across temporal, spatial and social scales (Hall & Leander, 2010). We chose family migration because the sharing of family histories benefits (Fivush, Bohanek, & Zaman, 2011), and migration continues to be a pressing issue globally, nationally, and in our communities. We hoped that this study would offer insight into how to facilitate learning and productive dialogue around such a timely matter.

In the current study, middle and high school youth (N = 17; self-identified as 6 male; 11 female; 13 African-American; 3 White; 1 Asian; Mean hours of attendance = 13; sample included 6 sibling pairs) created *family data storylines* to explore reasons for personal family mobility (What moved my family?) as well as national and global migration (What moves families?) in a free summer workshop at a city public library. Youth represented family decision-making and social conditions with online modeling and mapping tools and related the lives of their ancestors to their own experiences and futures. The three weekly workshop sessions (2 days per week, 5 hours per day) culminated in a public community exhibit.

Assembling family data storylines involved the following tasks: First, participants chose a side of the family to focus on. Second, participants chose one of two interactive web-based data tools accessed via their laptop's Internet browser: Social Explorer (Figure 1) or Gapminder (Figure 2). Both tools were selected because they afford (Gibson, 1979/1986) interactivity, as opposed to static data displays. Social Explorer is a historical thematic mapping tool that uses US demographic data. It accesses hundreds of variables from Census and other demographic datasets that go as far back as 1790. Data can be encoded as multiple visualization types—as dots, with colorful shading, with bubble size—and at national, regional, and local scales. In Social Explorer, users can also create different side-by-side temporal and spatial comparisons. Gapminder is a multivariable, dynamic graphing tool that uses public global socioeconomic data. It has five possible quantities or variables that can be selected. Y-axis, x-axis, color, and bubble size each have over 500 health and wealth indicators or measures to choose from. Timescales of datasets vary: Some start as early as 1800, and others only have one year of available data. In Gapminder, users can also alternate between logarithmic and linear scales, and select particular country bubbles to leave data trails over time. Third, participants selected variables from each tool's available datasets, like education level or household income in the US Census. Participants captured screenshots of models or maps and inserted them into a Microsoft PowerPoint, accompanied by slides or texts that explained data selections and what participants learned from the data.

### Learning environment design

Our learning environment design aligns with a 4E learning model. First, our design recognizes STEM modeling as distributed and embodied activities (Hall & Nemirovsky, 2012; Hutchins, 1995). We thus included activities beyond interacting with computers. For example, once or twice in each workshop session, the day started with a “four walls” game that set up comparisons between experiences of the youth and that of their ancestors, related to circumstances surrounding family migration (i.e., family size, educational attainment, occupation, neighborhood racial diversity). We asked questions such as: “How far along in school did your family member go?” (Wall 1: Grade school, Wall 2: High school, Wall 3: College and beyond, Wall 4: No formal schooling), which was followed by “How far along do YOU expect to go?” In Sessions 2 and 3, we performed a walking-scale timeline that, like the four walls game, asked participants to stand in for their family in historical time. These embodied comparison activities asked the participants to consider the historical social conditions that motivated or forced their families to move and that possibly could be explored with the data tools.

Second, we consider our design to be an example of extended learning with big data. Learning was distributed across workshop activities and family members. Youth assembled their storylines with siblings who were in the workshop and with parents, via phone calls, text messages, visits during the workshop day, and at-



home conversations over intervening nights and days, in which pieces of family story were filled in or changed or corrected by family members.

Third, our design embedded the sociotechnical practices of data science in a personal topic (family migration) and in a common, cultural, collaborative activity (family storytelling).

Finally, as described earlier, the focus for the current analysis is on enactive learning, or how the youth participants interacted with designed activities and context for storytelling and modeling with big data.

## Data collection and analysis

We video and audio recorded all activities and recorded participants' work on laptop computers with screen capture software. Our qualitative analysis took an interactionist approach (Greeno, 1994; Azevedo & Mann, 2018) towards engagements between individual agents and the learning environment. This approach, with a focus on both bodily interactions with the physical environment and tools as well as conceptual practices and perceptions of participants, supports the study of enactive and multimodal, embodied learning. We used interaction analysis methods (Jordan & Henderson, 1995) to understand participants' data wrangling strategies. As we reviewed participant trajectories (Jiang, 2018) across activities, we developed analytic memos around what appeared to be—through continuous or constant comparisons (Glaser, 1965) of individual participants—conceptual and technical practices to describe participants' data wrangling. We then selected episodes for microanalysis, in which we analyzed gesture, discourse, and tool use as sources of knowledge and learning. We often recreated maps or models in Gapminder or Social Explorer to better understand participant data explorations. The comparison of participants also focused on critical reflections on the data (i.e., data quality, data stakeholders), social history, and family migration. We reviewed and compared participant records until no new categories of data wrangling practices emerge (Strauss & Corbin, 1994). After identifying the categories of practices, we reanalyzed the selected episodes for microanalysis to trace the trajectories of practices and developed themes around the process of data wrangling.

## Findings

We define data wrangling as the practices that manage multiple datasets and measures in order to connect narratives with aggregate data. Data wrangling in an open data interface appears to involve the application of technical and statistical knowledge as well as an individual's social values and sense of identity. For instance, many of the participants' interactions with the data tools involved locating places that held personal meaning for them, whether in their family histories or own experiences, or selecting datasets that they identified with. In previous iterations, we have called this “getting personal with big data” (Kahn & Hall, 2016). Consequently, in our instructional context, data wrangling served as the means for participants to establish self–society relations (Kahn, under review).

### RQ1: What are the enactive practices that describe participants' data wrangling interactions with big data visualization tools?

We identified ten sociotechnical practices that comprise data wrangling for youth: 1) filtering data; 2) selecting indicators or variables; 3) data visual encoding; 4) interpreting data points; 5) identifying data patterns; 6) pursuing data surprises; 7) reasoning about data relationships; 8) countering data; 9) approximating data; and 10) making data predictions. These practices tended to be in the service of building comparisons to tell a story about both family history and broader socioeconomic trends (e.g., a comparison of origin place A to destination B that shows better economic conditions in B). Below, we introduce and define each practice and provide several brief illustrations.

In **filtering data**, participants either select a country bubble and all others fade in Gapminder or locate specific locations in a map in Social Explorer using the zoom feature (e.g., Figure 1). This practice results in highlighting specific data points or locations through a change of data visualization, but there is no change to the dataset being used to build the data visualization. Making such selections is challenging in an open data interface, when there are hundreds of datasets and variables to choose from. Consequently, personal connections tended to drive participant selections from the beginning, such as when youth selected a country in which their family members lived or zoomed into their neighborhoods on a map.



Figure 1. A participant is viewing the violent crime rate of her neighborhood in Social Explorer.

**Selecting indicators or variables** involves participants’ changing default indicators or variables by following specific constraints. This practice changes the dataset being used to build the data visualization. While the constraints guiding indicator selections varied among participants, the most common constraints included a) finding a dataset based on a predeveloped storyline, such as one participant’s selection of *Mean Years in School for Men 25 Years and Older* in Gapminder to explore whether her father moved to the U.S for better educational opportunities; b) following a scale, such as one participant’s successive selections of *Household Income Less Than \$10,000*, *Less Than \$50,000*, and *Less Than \$200,000* in Social Explorer; c) identifying oneself or people that participants were familiar with in the data, such as one participant (13 years old, female) selecting *Female Population Aged 10-14 Years* in Gapminder.

**Data visual encoding** is the practice of interpreting or understanding the relationship between two representations in which data is mapped into visual structures (e.g., color legends). For instance, in a choropleth map, areas are shaded in proportion to the measurement of variables (e.g., dark colors represent higher population density while light colors represent lower population density). Data visual encoding usually occurred when participants were trying to a) understand how the data is encoded in different visual representations, such as when one participant said, “Sure are a lot of crime rates” after hovering her mouse over a dark red area in Social Explorer; or b) change the details of visual encoding to leverage the data visualization for storytelling purposes. For example, in order to support the claim of moving due to safety concerns, one participant changed visual encoding from “one dot represents the value of 10,000” to “one dot represents the value of 5,000” in Social Explorer to make the differences in the number of crimes between two states more visible. As another example, guided by a researcher, one participant changed the x- and y-axis scales from linear to logarithmic in Gapminder to explore potentials of each visualization in explaining family story. Although the participant did not understand the statistical concept, she was aware of the change in visualization. A persistent challenge across interfaces for youth was differentiating between data encoding for rates or percentages and counts.

**Interpreting data points** refers to the practice of interpreting a single data point at one time. This practice was marked by highlighting with the cursor tooltip or selecting (filtering) a particular bubble in Gapminder or discrete geographic area in Social Explorer and by providing verbal explanations of the meaning of the numbers in the tooltip in discourse, as illustrated below.

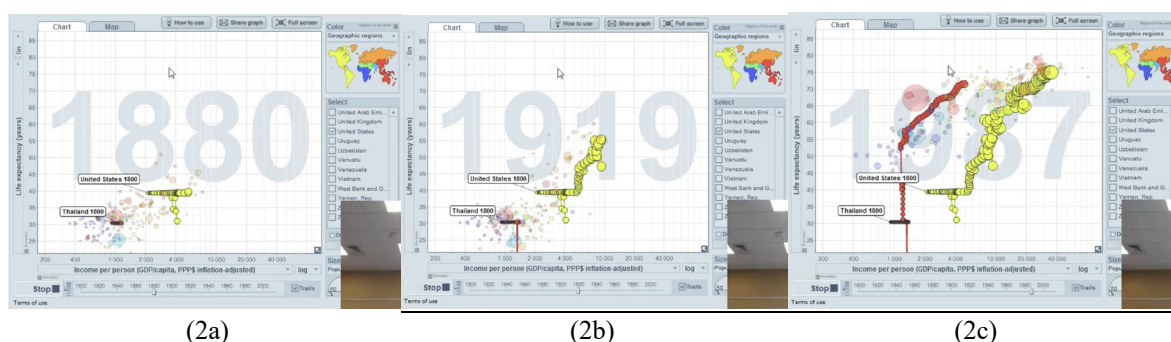
Excerpt 1 [Week 1, Day 1]

Francine (*hovers mouse over US bubble in Gapminder, turns to her tablemate*): See, right here. It's United States. It's about how much money they earn, 53.4, and how long they live, 79.1. (1)

**Identifying data patterns** represents the practice of describing a data trend, such as the trend of a variable for a single geographic area over time (e.g., for a state 1920–1970) or the trend of a variable for multiple areas at one point in time (e.g., a regional trend in the year 2000). For instance, in the excerpt below, participant Sage, whose mother came from Thailand and father was born in the US, described a data pattern for life expectancy while watching an animation in Gapminder in which she compared the two countries.

Excerpt 2 [Week 1, Day 2]

Sage: It's going down a little bit for the United States (*Figure 2a*).  
 Sage (*while seeing a drop in life expectancy in Thailand and in the US in 1919 [Figure 2b]*): Oh, no!  
 Sage: Life expectancy is going up quite a bit in both countries now (*Figure 2c*).



**Figure 2a, 2b, 2c.** In Sage's Gapminder model, the US and Thailand are selected, and time plays from 1800 through 2015. Life expectancy (y-axis) and income (x-axis) generally increased over time for the US and Thailand, except for a dramatic drop in life expectancy in both countries in 1919 (2b).

**Pursuing data surprises** involves participants' actively looking for or noticing outlier data, typically represented by a divergent bubble trail in Gapminder or a very light or dark region in Social Explorer. For instance, while noticing the drop in Figure 2b, Sage went to search online for possible reasons for a dramatic decrease in life expectancy in Thailand in 1919. This example indicates the close relationship between the practices of identifying data patterns and pursuing data surprises.

**Reasoning about data relationships** refers to when participants describe relationships between two or more variables or covariation to explain family migration. In most cases, participants had difficulties relating two variables and tended to examine indicators (e.g., household income, educational attainment) independently.

**Countering data** involves sharing a personal experience or account that does not align with the data. For example, Francine explained that her father went to school every year while noticing that the average number of years of schooling for men in Sudan, according to Gapminder, was only 3 years in the 1980s. Another participant, Naimah, who assembled a Social Explorer comparison showing an increase in the Black population in the North and a decrease in Black population in the South between 1920 and 1970, noted that Social Explorer did not include how "moving from South Carolina affected [her] family's connections, and how it was a milestone in [her] history." We view this kind of critique as a challenge to what the Census takers deem as important for understanding social history. This practice shows that participants had agency and authority in questioning big data, which challenges the contention that personal experience serves to confirm data for youth (Enyedy & Mukhopadhyay, 2007).

**Approximating data** describes when participants make selections that are approximations for the data they were looking for when there is no data for the exact year, category, indicator, country or place that they desire. For example, for Naimah, survey years 1920 and 1970 roughly corresponded with the birth of her paternal great-grandparents in South Carolina and their decision to leave (1920) and her grandparents' return to the South (Alabama) from Illinois (1970). These approximations can be understood as emergent fixes to manage uncertainty (Star, 1985) or missing data in data wrangling.

**Making data predictions** represents the practice of imagining what the data could show in the future. In our analysis, this practice did not occur while participants interacted with Gapminder and Social Explorer, although participants considered their own futures in our activities. For instance, in her family data storyline, Hannah compared the percentage of persons 25 years and over with some college experience or more between Chicago and Oakville in 1990 to examine whether her father moved to Oakville for college. She subsequently included a slide in her family storyline describing where she wanted to attend college in the future.

## RQ2: What is the conceptual process of data wrangling to assemble family data storylines?

We developed a conceptual model (Figure 3) of the data wrangling process that contains four interrelated recursive stages: 1) *finding my family and me in big data*, 2) *understanding my family and me in relation to the society*, 3) *challenging big data with authority and agency*, and 4) *constructing a big data model to share my family story*. During the stage of *finding my family and me in big data*, participants selected data and indicators related to themselves. They described how historical events affected family migration in the stage of *understanding my family and me in relation to the society*. In the stage of *challenging big data with authority and agency*, participants critically analyzed data when the data was not consistent with their initial understandings of family stories and actively explored alternative indicators that might contribute to family migration. They selected indicators and visualizations that aligned with the family story in the stage of *constructing a big data model to share my family story*. Participants tended to start with finding data related to themselves in this process (*finding my family and me in big data*), which is a “first move” in data wrangling, and ended with constructing models for assembling their family data storylines (*constructing a big data model to share my family story*), which they presented to their peers and were displayed in a public exhibit.

The four stages are recursive, and participants moved back and forth between these stages. Although participants’ conceptual movements between these stages indicate a variety of paths in the process of data wrangling, some movements were more challenging for the youth. For instance, Sage engaged in comparing different variables (e.g., income) between Thailand and the U.S in the stage of *finding my family and me in big data*. Her mother moved from Thailand to the U.S for love, which is an abstract concept that she had difficulties in finding variables to represent in the stage of *constructing a big data model to share my family story*. In her case, the movement between these two stages rarely occurred after several failures in finding appropriate variables to represent the abstract concept.

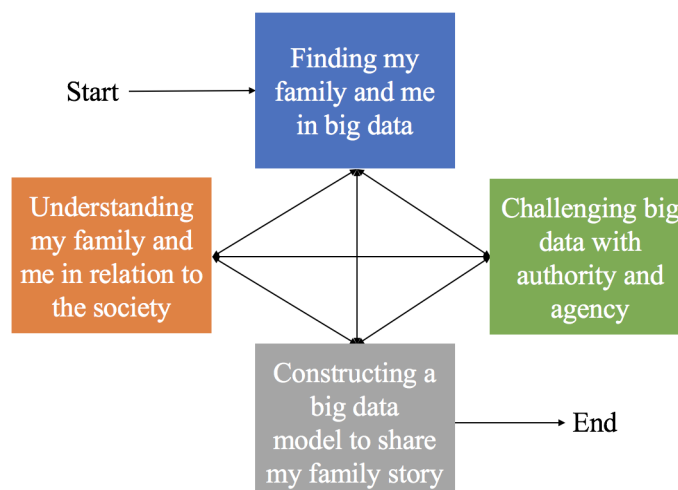


Figure 3. A conceptual model of data wrangling to assemble family data storylines.

## Discussion and implication

Our study illustrates what data wrangling looks like in an informal learning setting and how students engaged with various sociotechnical practices in the process. Participants engaged in learning about their families to learning about society through data wrangling. Typically, one practice went hand-in-hand with other practices (e.g., interpreting the meaning of a dark red county involves both the practices of visual data encoding and interpreting data points) in interrelated recursive stages.

The study shows that the “first move” in data wrangling tends to be finding data related to oneself while youth navigate data visualization tools to assemble their family storylines. This finding supports the current understanding of the youth’s interactions with big data while contributing new insights: Whereas research in data science and STEM education has evidenced the value and challenges of culturally relevant pedagogy (e.g., Enyedy & Mukhopadhyay, 2007), this study illuminates that getting personal with big data provides an entry point and access to interactions with big data technologies. While other instructional designs with these tools reduced the number of indicators available for exploration (e.g., Radinsky et al., 2014), we found that it is important to provide a rich and inclusive dataset to offer equitable learning opportunities in data

wrangling. This finding suggests that future designs could engage youth in exploring big data technologies with intentional scaffolds for getting personal with data.

Also, this study addresses tensions of confirming data presented in Enyedy and Mukhopadhyay's (2007) study, in which youth use data to confirm pre-existing conclusions without questioning the model. Our study demonstrates that family data storylines fostered the youth's authority and agency over big data, and this produced both questions about and challenges to the big data. This finding has implications for designing learning activities with big data in which youth have their own voices. The literature supports this design principle and studies have shown the importance of building a close relation between the self and digital artifacts (e.g., Philip et al., 2016). Future research can pursue other personal contexts for big data exploration, such as daily mobility datasets or personal health data.

In addition, our study presents several challenges that need to be addressed in data wrangling practices in future studies. The first challenge is ensuring the understanding of statistical concepts, such as the difference between counts and rates, as well as the understanding of how to make a reasonable comparison, such as zooming two maps to the same level to set up comparisons. A second challenge is engaging youth in discussions of whether variables in a dataset can be used to measure abstract concepts (e.g., love). A third challenge is encouraging more data exploration when the model does not correspond with family stories. In one such case, data wrangling consequently went unresolved (i.e., no data was included in Isis's final family data storylines).

Although each youth assembled their own family data storyline on an individual laptop, participants frequently interacted with tablemates, who were often their siblings, when they discovered data that they were personally connected to, such as data representing the city where participants lived. Future research is needed to investigate how to support getting personal with data in more collaborative learning settings, such as data modeling activities designed for intergenerational families or teams of peers. In addition, this study shows that youth had more exchanges with each other when they encountered trouble in data wrangling, such as when they encountered data that did not align with their initial understandings of why their families moved. These exchanges, particularly between siblings, entailed negotiations of authority in interpreting data and telling the family history, such as Isis told her older sister Naimah, "I do not need you to tell me about my father's story." Much more needs to be learned about the effect of agency and authority on peer interactions with big data.

Our findings regarding data wrangling practices and process should be the starting point in research on designs for learning and instruction that promote modeling narratives with big data visualization technologies in different contexts. For example, future studies can examine whether some practices (e.g., identifying data patterns) are more influenced by the tool affordances (e.g., bubble trail in Gapminder) while some practices (e.g., pursuing data surprises) are more universal across data wrangling activity and interfaces. Although these practices were used with two interfaces (motion charts and web maps), we expect they would apply to other data interfaces involving multimodal, multivariable dynamic representations (e.g., stacked graph and radar chart). We believe that big data technologies are valuable instructional tools for offering learning opportunities and communication with others about important social and scientific issues.

## Endnotes

(1) All names are pseudonyms.

## References

- Azevedo, F. S., & Mann, M. J. (2018). Seeing in the dark: Embodied cognition in amateur astronomy practice. *Journal of the Learning Sciences, 27*(1), 89-136.
- Bandura, A. (1986) *Social foundations of thought and action: A social-cognitive view*. Englewood Cliffs, NJ: Prentice-Hall.
- Borner, K., Peppler, K., Kennedy, B., Uzzo, S., and Heimlich, J. (2017). "Data Visualization Literacy: Research and Tools That Advance Public Understanding of Scientific Data" *NSF - National Science Foundation* Available at [https://www.nsf.gov/awardsearch/showAward?AWD\\_ID=1713567](https://www.nsf.gov/awardsearch/showAward?AWD_ID=1713567)
- boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society, 15*(5), 662-679.
- Engel, J. (2017). Statistical literacy for active citizenship: A call for data science education. *Statistics Education Research Journal, 16*(1), 44-49.
- Enyedy, N., & Mukhopadhyay, S. (2007). They don't show nothing I didn't know: Emergent tensions between culturally relevant pedagogy and mathematics pedagogy. *The Journal of the Learning Sciences, 16*(2), 139-174.
- Fivush, R., Bohanek, J. G., & Zaman, W. (2011). Personal and intergenerational narratives in relation to adolescents' well-being. *New directions for child and adolescent development, 2011*(131), 45-57.

- Gibson, J. J. (1986). *The ecological approach to visual perception*. Hillsdale, NJ: Erlbaum. (Original work published 1979)
- Glaser, B. G. (1965). The constant comparative method of qualitative analysis. *Social problems*, 12(4), 436-445.
- Greeno, J. G. (1994). Gibson's affordances. *Psychological Review*, 101, 336-342
- Greeno, J. G., & Engeström, Y. (2014). Learning in activity. In K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 128-147) (2nd ed.). London: Cambridge University Press.
- Hall, R. & Leander, K. (2010, July). Scaling practices of spatial analysis and modeling. In Hall, R. (Symposium Chair). Scaling practices of spatial analysis and modeling. International Conference of the Learning Sciences.
- Hall, R., & Nemirovsky, R. (2012). Introduction to the special issue: Modalities of body engagement in mathematical activity and learning. *Journal of the Learning Sciences*, 21(2), 207-215.
- Harrison, L., Yang, F., Franconeri, S., & Chang, R. (2014). Ranking Visualizations of Correlation Using Weber's Law. *IEEE Trans. Vis. Comput. Graph.*, 20(12), 1943-1952.
- Hutchins, E. (1995). How a cockpit remembers its speeds. *Cognitive Science*, 19(3), 265-288.
- Jiang, S. (2018). STEM+ L: Investigating Adolescents' Participation Trajectories in a Collaborative Multimodal Composing Environment. *ProQuest LLC*.
- Jordan, B., & Henderson, A. (1995). Interaction analysis: Foundations and practice. *Journal of the Learning Sciences*, 4(1), 39-103.
- Kahn, J. (under review). Learning at the intersection of self and society: The family geobiography as a context for data science education. *Journal of the Learning Sciences Special Issue: Situating Data Science*.
- Kahn, J., & Hall, R. (2016, April). Getting personal with big data: Stories with multivariable models about global health and wealth. Paper presented at the American Education Research Association 2016 Annual Meeting, Washington D.C.. Available in the AERA Online Paper Repository.
- Kosara, R., & Mackinlay, J. (2013). Storytelling: The next step for visualization. *Computer*, 46(5), 44-50.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, England: Cambridge University Press.
- Philip, T. M., Olivares-Pasillas, M. C., & Rocha, J. (2016). Becoming racially literate about data and data-literate about race: Data visualizations in the classroom as a site of racial-ideological micro-contestations. *Cognition and Instruction*, 34(4), 361-388.
- Philip, T. M., Schuler-Brown, S., & Way, W. (2013). A framework for learning about big data with mobile technologies for democratic participation: Possibilities, limitations, and unanticipated obstacles. *Technology, Knowledge and Learning*, 18(3), 103-120.
- Polman, J. L., & Hope, J. M. (2014). Science news stories as boundary objects affecting engagement with science. *Journal of Research in Science Teaching*, 51(3), 315-341.
- Radinsky, J., Hospelhorn, E., Melendez, J. W., Riel, J., & Washington, S. (2014). Teaching American migrations with GIS census webmaps: A modified "backwards design" approach in middle-school and college classrooms. *Journal of Social Studies Research*, 38(3), 143-158.
- Rubel, L. H., Hall-Wieckert, M., & Lim, V. Y. (2017). Making Space for Place: Mapping Tools and Practices to Teach for Spatial Justice. *Journal of the Learning Sciences*, 26(4), 643-687.
- Star, S. L. (1985). Scientific work and uncertainty. *Social studies of Science*, 15(3), 391-427.

# Promoting and Tracing High School Students' Identity Change in an Augmented Virtual Learning Environment

Hamideh Talafian, Mamta Shah, Amanda Barany, and Aroutis Foster  
Ht343@drexel.edu, mamta@drexel.edu, amb595@drexel.edu, aroutis@drexel.edu  
School of Education, Drexel University

**Abstract:** This paper illustrates the application of Projective Reflection (PR) as a theoretical framework in an ongoing NSF-CAREER study to trace learning as identity change among 20 high school students. A 9-week augmented virtual learning environment (AVLE) course, *Virtual City Planning*, was created and implemented in a science museum classroom. In-game and in-class student data and mentor observations were collected and analyzed using Quantitative Ethnography (QE) techniques to identify patterns of identity exploration enacted over time. This paper offers group findings to illustrate general trends of identity exploration during the AVLE, and one illustrative case study to provide an in-depth look at an integrated identity exploration trajectory. This work concludes with a discussion of the need for promoting identity exploration as an educational goal, and a means to support learners to adapt to the changing workforce of tomorrow. The affordances of AVLEs are discussed to advance research in this nascent area.

## Introduction

Educational research has increasingly emphasized ways to develop learner skills in *identity exploration*, or the intentional repeated examination of who a learner is and who he/she wants to become (Kaplan & Flum, 2012), which can lead to identity change in targeted directions over time (i.e. change towards a future in a STEM career) (Foster, 2014). Virtual learning environments (VLEs) such as games have been increasingly identified and researched for their potential to support change in both cognitive (i.e. knowledge) and affective (i.e. motivation) elements of self (e.g. Kamarainen, Metcalf, Grotzer, & Dede, 2015). To rigorously leverage these affordances, however, research points to the need for robust conceptual and design frameworks for developing game environments (DeVane, 2010), and the purposeful inclusion of real-world pedagogy that can supplement game content as part of an integrated learning experience (Clark, Tanner-Smith, & Killingsworth, 2015). We refer to this merging of a primarily virtual environment with real-world augmentations as an augmented virtual learning environment (AVLE) (Milgram, Takemura, Utsumi & Kishino, 1994), which integrates the benefits of both worlds to comprehensively provide identity exploration opportunities. Rowsell & Abrams' (2011) I/identity framework offers a useful starting point with its "multidisciplinary approach to mediating I/identity in the 21<sup>st</sup> century," with consideration for both individual (i)identity change, or "an embedded sense of self that inherently fuels and/or shapes one's behavior" (p. 4), and collaborative aspects of (I)identity change that are "rooted in practices and overt forms of meaning-making" (p. 4) offered by AVLEs.

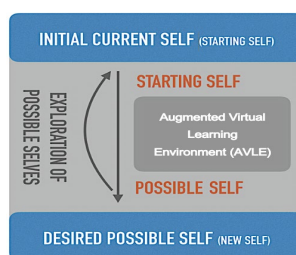
Thus, in this paper, we illustrate one way of conceptualizing learning as a process of exploring possible role selves through experiences in AVLEs to gain the knowledge and skills to reconstruct oneself through a theory called as Projective Reflection (Foster, 2014). The paper describes the application of Projective Reflection in an ongoing NSF-CAREER study that informed the design and implementation of *Philadelphia Land Science (PLS)*, a text and web-based VLE for exploring environmental science and urban planning roles. The in-game experiences mediated by PLS were augmented by an external in-class curriculum to further scaffold students' exploration of the two roles. Thus, the combination of PLS + the external curriculum was referred to as the Virtual City Planning AVLE, which was offered to 20 high school students over a 9-week period in a science museum in a northeastern US city. Data corpus included in-game logged data, research memos, pre-post interview questions, written and visual artifacts created in-game and during external individual and group curricular, reflection and discussion activities. These data sources were used to examine the following research question: "To what extent did participants engage in an integrated identity exploration in Virtual City Planning as a result of exploring the roles of an environmental scientist and urban planning?" Implications are discussed to advance educational research on framing, facilitating and tracking learning and identity in AVLEs through the use of emerging frameworks such as Projective Reflection.

## Theoretical framework

Projective Reflection (PR) is a theoretical framework that conceptualizes learning as identity exploration that facilitates learners' engagement in self-transformation or identity change in AVLEs and explicates the process of



learning as identity exploration over time (See Figure 1) (Foster, 2014). Four theoretical constructs in PR support identity exploration, and structure assessments of learners' cognitive and affective change over time: (1) knowledge, (2) interests and valuing, (3) self-organization and self-control, and (4) self-perception and self-definition (Foster, Shah, & Barany, 2017). Six questions are used to synthesize these changes into a comprehensive characterizations of a learner's initial current self, their exploration of possible selves (measured repeatedly across the experience), and the new self at the end of the intervention (Foster, 2014): 1) what the learner knows – current knowledge, 2) what the learner cares about – self and interest/valuing, 3) what/who the learner expects to be throughout the virtual experience and their long term-future self, 4) what the learner wants to be – possible self, 5) how the learner thinks – self and interest, and 6) how the learner sees him/herself – self-perception and self-definition (see Table 1).



**Figure 1.** The Projective Reflection framework for conceptualizing learning as a process of identity change over time.

Over the duration of an intervention, a learner is encouraged to engage in targeted and intentional reflection on aspects of self along the four PR constructs, including (a) her/his starting self at the beginning of a learning experience, (b) repeatedly throughout the exploration of possible selves as supported by the AVLE, and (c) her/his new self at the end of the intervention. Furthermore, AVLEs are designed to have opportunities for participants to play, engage in curricular activities, reflect and discuss in relation to the roles explored both in the game and outside (Foster & Shah, 2016). For each student, the process of tracing change is carried out chronologically, guided by the six questions which allow us to explicate the extent to which the learner explored a possible role by developing intentional changes in their knowledge, interest and valuing, self-perception and self-definition, and self-organization and control skills- indicating the extent to which the identity exploration was comprehensive or integrated. In this paper, we report the case of a female student as an elucidation of one integrated identity exploration trajectory over the 9-weeks from the start of her participation in Virtual City Planning (weeks 1-2), during the course (weeks 3-7), and the end of the course (weeks 8-9). Group patterns of identity exploration are also identified to not only better understand the case study, but also to highlight the overall changes in the four constructs informing identity exploration as a result of engaging in the AVLE over the 9 weeks.

## Methodology

This research was conducted as part of an ongoing 5-year NSF CAREER project awarded to advance theory and research on promoting identity exploration and change in science using virtual learning environments using Projective Reflection (Foster, 2014). Building on this broader agenda, researchers designed and implemented an AVLE that featured the weekly use of the VLE *Philadelphia Land Science (PLS)* and supportive real-world augmentation in the classroom (for more information see Shah et al., 2018). The resulting AVLE, titled Virtual City Planning was offered to a total of 54 Philadelphia high school students who participated in weekly STEM career programming at their local science museum. *Virtual City Planning* was offered to facilitate learning as identity change by engaging participating students in real-world activities on environmental science and urban planning. *Philadelphia Land Science (PLS)* allowed players to explore roles related to urban planning and environmental science careers as they connect to the Philadelphia context and redesigned/rezoned the map of the city with different stakeholders. In *PLS*, groups of students were synchronously guided by online and in-person mentors to collaborate with their peers in redesigning the map of the city. Building on the strengths of *PLS*, the research team designed real-world features (i.e. non-digital design and reflection activities) that leveraged affordances of classroom context to further support identity exploration.

Three versions of Virtual City Planning were offered from September, 2016 to May, 2017, each successively refined from the previous offering (9-weeks in Fall with 20 students, 8-weeks in Winter with 19 students, and 4-weeks in Spring with 19 students) to strengthen the extent to which the activities in the AVLE could support students' process of projective reflection (Cobb, Confrey, DiSessa, Lehrer, & Schauble, 2003). This paper focuses on the first iteration of Virtual City Planning, which was enacted over nine weekly 90-minute



sessions with 20 students in the science museum. 60% of them were females (n = 12), 30% were males (n = 6) and the other 10% preferred not to reveal their sex (n = 2). They attended a local magnet school which focused on STEM preparation and had several programs for teachers' professional development and students' STEM enrichment integrated with/within the museum.

## Procedures and data collection

Data sources included logged in-game data, classroom artifacts, and researcher observations. Once collected, data was organized chronologically by student to track PR changes from the beginning, at repeated points during, and at the end of the AVLE.

Over the 9-weeks, Virtual City Planning allowed participants to learn about the process of proposing a rezoning proposal for the city of Philadelphia that addresses the competing and complementary needs of four stakeholders in a well-balanced manner. Activities in Philadelphia Land Science and outside of it were designed to facilitate intentional shifts in what learners know, how they think, what they care about, how they see themselves, what they want to be and expect to be in relation to environmental planning and urban science. For the first half of the course, learners were assigned to one of the four stakeholder groups to explore specific environmental and/or economic needs of the city. The four stakeholders were Philadelphia Institute for Environment Preservation, Philadelphia Economic Affairs Coalition, Environmental Council for Greater Philadelphia, and Bridgeway Community Action Group. In the second half of the course, students were jigsawed in a manner that every new group had a representative from each of the four stakeholder groups; thus, simulating the real-world scenario of multiple interests and expertise coming together to address real-world issues. The two parts of the course prepared students towards the goal of proposing a rezoning plan.

In Session 1, we engaged players in learning about their teams and understanding the expected workflow of PLS. In this session, we also gathered data to establish students' starting self by way of an intake interview in PLS, background survey, and a focus group discussion. Sessions 2 and 3 required participants to research the stakeholders' concerns from various factions of the city (commercial, residential, environmental, civic, social) and understand what is expected of the urban scientists in the 'request for a re-zoning proposal.' Sessions 4 and 5 engaged players in creating models to create and test the extent to which their re-zoning proposal balanced the needs of the stakeholders on an interactive map of Philadelphia using a mapping tool called iPlan. They received feedback from the stakeholders and their mentors which lead to iterative refinements in sessions 6-8. Data were obtained from sessions by way of in-game logged data (chats, notebook entries, iplan maps, researcher memos) to understand what each participant was focusing on more or less along the four constructs as he/she was exploring the role assigned and those that emerged naturally, and how the participant was engaged in the identity exploration process individually, in participation with PLS, and within the classroom setting (interaction with peers, mentors). Finally, in session 9, players worked towards finalizing a written document explaining the rezoning plan and representing the plan on a map. Data was collected to establish participants new self by way of an in-game exit interview, and a focus group discussion.

## Data analysis

Data was coded both inductively and deductively to answer the question of "To what extent did participants engage in an integrated identity exploration in Virtual City Planning as a result of exploring the roles of an environmental scientist and urban planning?" Overall, Quantitative Ethnography (Shaffer, 2017) was adopted to guide the data analyses procedures. First, researchers engaged in a deductive coding process for each case (Krippendorff, 2004) using the qualitative analysis software MAXQDA. Lines of student data were coded as illustrating self-reflection on demonstration of the four PR constructs, with the agreement reached by two graduate-level coders. Thereafter, qualitatively coded chronological data was quantified using a binary approach. Each line was coded for the occurrence (=1) or non-occurrence (= 0) of four constructs in order to prepare the data for Epistemic Network Analysis. Epistemic Network Analysis (ENA) (Shaffer, 2017), a tool that can be used with Quantitative Ethnography, was employed for modeling the structure of connections in the data to look for discernable patterns between the identity of participating students demonstrated at the start, during, and at the end of their participation in Virtual City Planning. This association structure between the changes in their knowledge, interest and valuing, self-organization and self-control and self-perception and self-definition, was modeled mathematically by creating an adjacency matrix based on their co-occurrence in the course over time, by the three data points. As such, ENA offered a unique way to recognize the patterns of identity exploration at both the group (Big I) and individual levels (Little i) for engaging in projective reflection as a result of Virtual City Planning. We referred back to the interactions and activities coded in the data to close the interpretive loop and thus fully understand the phenomenon mirrored in the model for each student and the group at large.

## Findings

Group findings are presented in this section as the themes across four constructs of PR: knowledge, interest and valuing, self-organization and self-control, and self-perception and self-definition. Then, individual findings for one case “Andrea” will be illustrated by Epistemic Network Analysis (ENA) and explained.

### Group findings

#### Change across knowledge

Based on the results of the inductive-deductive qualitative coding of student data, three thematic patterns related to student knowledge changes were identified: (1) demonstrations of core content knowledge and problem-solving and critical thinking, (2) shifts from general to specific knowledge demonstrations over time, and (3) connecting new information to existing awareness of situated context.

*Demonstrations of core content knowledge and problem-solving and critical thinking.* The majority of codes applied to student data identified instances in which students demonstrated core content knowledge gains, such as an understanding of what urban planning is and how it is conducted. These trends exist largely due to the design of the AVLE. For example, the intake survey prompted James to describe what urban planners do, to which he responded: “They collect data about the places around us and get important people to change it.” Similarly, as a result of the design of the AVLE, a large proportion of codes identified instances in which learners engaged in problem-solving and critical thinking by generating solutions to in-game problems, or by offering justifications for their design choices. For example, Alice explained her choice to add more industrial zoning to downtown Philadelphia, saying “people need more places to work and find jobs to work” which supports the affordances of AVLEs for developing meta-knowledge skills.

*Shifts from general to specific knowledge demonstrations over time.* A review of shifts in knowledge demonstrations over time within individuals reveals a general shift from broad understandings of urban planning and environmental science towards more specific recognition of the complex relationships between domain variables over time. For example, changes in Kimberly’s responses to the question “What factors might an urban planner need to consider when planning/designing a city?” reveal increased specificity and a deeper understanding of urban planning processes from engagement in the AVLE. In week 1, she wrote that urban planners “need to factor in the needs of other people” while in week 9 she was able to explain how “[urban] planners must think and fit the needs of stakeholder. Stakeholders focus on things they like to see or not like to see for specific cases.”

*Connecting new information to existing awareness of situated context.* As students demonstrated increases in urban planning and environmental science over time, they also regularly connected the urban planning problems, processes, and AVLE context (Philadelphia) to the knowledge of similar issues as they experience them in their own communities. For example, as Ali encountered the economic and environmental issues emphasized in the AVLE, he was able to reflect on the issues he had noticed in his own neighborhood (pedestrian safety). He later applied his new knowledge of how to identify problems and design solutions to his own context: “if I would want to change things, I would put more safety signs on streets to make people feel safe. my changes may help because driving lately has been more of a problem because people are frightened of others on either the highway or just a simple street.” These connections demonstrate the potential of the augmented AVLE for promoting deeper learning integrated into learners’ understandings of their own interest and values, patterns of self-organization and self-control, and self-perception and self-definition.

#### Change across interest and valuing

After the analyses of qualitative data for interest and valuing, two themes emerged: (1) change from personal interest to less personal goals and interests, (2) overlapping changes in interest and knowledge.

*Change from personal interest to less personal goals and interest.* The students’ interest and valuing changes showed a significant difference from the beginning to the end of the AVLE. Yet, the direction of the change from more personal goals and interests in the beginning weeks to less personal interest was an evident trend in all students. Deductively, many of the students’ reflections and responses to survey questions were coded personal interest in initial weeks, while similar questions toward the end of the AVLE led to more “global relevance codes” which demonstrates a shift to less personal goals when the player recognizes how tools and content related to a community or society and satisfies the needs of the group members. For example, when Andrea was prompted with the question of “How the design of Philadelphia has affected you?” in post-survey she responded: “... [environmental issues] can impact us by changing the city it can make a better environment. If we change the ecosystem to be better, it can make Philly look good and it will be healthier for us” which shows a well-suited response for “global relevance” as well as “personal relevance”. This was just one example of many responses demonstrating students’ less personal interests toward the end of the AVLE.

*Overlapping changes in interest and knowledge.* Another evident theme in students' responses regarding interest and valuing was that change in this construct happened in concert with a change in knowledge, they demonstrated their interest using new words and expressions which were inductively coded as knowledge, too. For example, when Ellen was asked to report her team's recommended value of bluebirds which are among endangered species in the Northeast, she answered "The recommended value of bluebirds is 2270 bluebirds. The stakeholders believe this is the perfect amount because if there were more there would be too much environmental space and not as much housing. However, if there were less than the city would not have the environmental space that it needs." A justification which shows Ellen both has the knowledge of keeping endangered species in her surrounding environment and values people's needs in having environmental space they need to.

#### Change across self-organization and self-control

The emerged themes came out of the researchers' memos from the students' demonstration of self-organization and self-control: (1) Asking help from the more knowledgeable peers/stakeholder, (2) Working toward a common goal and demonstrating metaknowledge.

*Asking help from more knowledgeable peers/stakeholders.* Asking help from the knowledgeable peers/stakeholders came mainly from the researchers' memos recording instances when students were asking for help when discussing their final map changes with their peers and stakeholders which was coded co-regulated learning. In co-regulated learning examples, the students were asking for help from more knowledgeable peers, stakeholder, and sometimes instructors.

*Working toward a common goal and demonstrating metaknowledge.* When the students were working in groups toward a common goal in addressing the stakeholders' concerns, they were not only demonstrating socially-shared regulated learning but also problem solving, critical thinking, communication and collaboration skills which were meta-knowledge components under knowledge construct. For example, when they were asked to justify the rezoning of their map "...[zoning for business] is important because there is not a lot of space for business so we are thinking to make more space for it." First, they indicated that they are looking for a way to solve the issue (problem-solving and critical thinking), second, they were indicating that they are working toward a common goal by bringing a plural pronoun (we). This example with many other examples of this sort informed us that the students worked toward the common goal with their stakeholder, and brainstormed ideas with the members of the group to resolve the issues and rezone the maps.

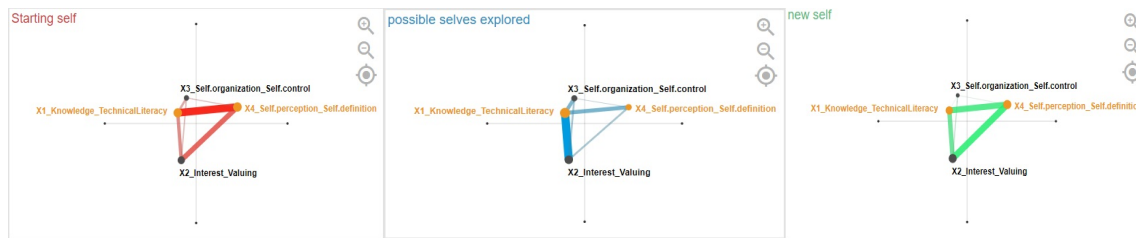
#### Change across self-perception and self-definition

The change in this construct was traced by coding self-efficacy, self-concept, and possible self-explored as the three main sub-constructs under it which led to one overarching theme: (1) No eloquent change in students' self-perception and self-definition.

*No eloquent change in students' self-perception and self-definition.* The analysis of students' qualitative data did not show drastic changes in students' self-perceptions and self-definition. Nearly all the students responded the same to the prompt of "what do you want to be in the future? what do you NOT want to be in the future?" before and after the AVLE. For instance, Ali said "basketball player" Ciara said "wealthy, happy, educated" Elijah said "business owner" and many other similar responses which were not changed in the post-survey. This indicates that the students' perceptions of themselves, their interest and their future desires barely change.

### Case findings

The illustrative case example Andrea demonstrated intentional reflection and integrated change in all four areas: knowledge, interests and values, patterns of self-organization and self-control, and perceptions and definitions of self over time in three levels of starting self, possible selves explored, and new self which is summarized Table 1. Figures 3-6 illustrate the strength of PR connections in these levels and the strength is calculated from the number of coded segments using ENA. Overall, Andrea demonstrated many instances of foundational knowledge in relation to self-perception and self-definition with the hope of working in the science field and "contribut[ing] to some type of urban planning" in initial weeks of using AVLE (Figure 3). Later in the course, Andrea's identity exploration focused more on the interest and valuing aspects as a result of engaging in the AVLE; this was integrated her emergent and detailed knowledge of urban planning (see Figure 4). Finally, a good balance between her knowledge, interest and valuing, and self-perceptions and self-definitions were revealed which is also shown in Figure 4.



**Figure 3.** Andrea’s Starting Self. The connection between knowledge and self-perception and self-definition is the

**Figure 4.** Andrea’s Exploring Possible self. The connection between knowledge and interest

**Figure 5.** Andrea’s new self, shows a good balance between her knowledge, interest and valuing, and self-perception and self-definition.

**Andrea’s starting self**

The adjacency matrix of knowledge, interest and valuing, and self-perception and self-definition in Andrea’s starting self (Figure 3) during the first two weeks of the course reveal that the connection between knowledge and self-perception and self-definition constructs are stronger than others. This strength comes from her “Agree” responses to Likert-scale questions which were assessing both knowledge and self-perception and self-definitions constructs such as: “I would be able to express my view in form of a group of people” or “I would be able to create a plan that addresses the issue”. Also, the connection of these constructs with interest and valuing construct is not as week as their connection to self-organization and self-control. Andrea demonstrated interest and valuing by describing her interest in the environment “I want to help make the environment we live in clean” understanding the relevance of designing the city to her personal life “the design [of the city] affects me because of too many buildings and it gets so crowded”. She also proposed suggestions for rezoning the city by telling “we should get rid of some buildings or separate them into different areas of Philadelphia, so we can have enough room and make it feel less crowded” which demonstrated her problem-solving abilities in relation to relevance during the first two weeks of the course.

**Andrea’s exploring possible self**

Figure 4 shows the adjacency matrix for Andrea’s exploring possible self which illustrates her change from week 3-7. Comparing to her starting self, Andrea demonstrated more knowledge and interest and valuing construct, and her adjacency matrix showed weaker connection to other constructs. Totally, knowledge and interest and valuing constructs together in all activities were coded (n=205) times which obviously have a bigger portion in Figure 4. For knowledge, she demonstrated increasingly detailed knowledge of urban planning by suggesting her detailed recommended land parcel changes “We changed the industrial areas to open wetlands so that the wetlands couldn't be polluted. WE decided that the open wetlands should be just on the wet side of the map, near the water so that the lands or the people around the area didn't pollute it and poison any of the animals or the creatures on that part of the map/area.” She capitalized “WE” which means that she is emphasizing her collaboration with other team members and socially shared learning. She also demonstrated many instances of relevance which made her adjacency matrix thicker in this construct, too: “Me as a citizen has to show how people should respect how to treat the environment. I think that our changes will affect the area in a good way and in a bad way because people will need to find more industrial areas and it will be good because the environment will be safe.” However, the link between knowledge and self-perception and self-definition is weaker as well as self-perception and self-definition and interest and valuing comparing to Figure 3 (her starting self). One explanation may be that she had a better understanding of each of these tasks because the AVLE had simulated real-life activities and allowed her to self-reflect during the program, thus, picked a less direct assertive response to the questions that were assessing those constructs.

**Andrea’s new self**

Figure 5 illustrates Andrea’s adjacency matrix for her new self in final weeks of 8-9 which shows a good balance between her knowledge, interest and valuing, and self- perception and self-definition. This figure is well supported by qualitative themes as we discussed earlier “overlapping changes in interest and knowledge”. For the majority of students, many instances of knowledge co-occurred to the interest and valuing utterances and in some cases, they were both presents in a single utterance. For example, when Andrea was asked “Whose needs and opinions should a city plan reflect?” she responded, “it should reflect of the citizens of United States of America” compared to some other peers who had answered “people” or “citizens of the city”. This is a detailed utterance which both demonstrates a good understanding of the city planning goals and the needs of the people who live in a country asserting the fact that any decision in rezoning the city will impact the whole population of America indirectly.

This utterance was coded both for core content foundational knowledge and global relevance which was a sub-category of interest and valuing.

Analyzing Andrea’s adjacency matrices from starting self to new self, and the similarity of starting self to new self, uncovered the fact that Andrea’s responses to many questions of knowledge and self-perception and self-definition were unrealistic at the beginning of the program. Using AVLE, she could explore her identity and challenged some of her notions about her capabilities (see Figure 4). However, by using AVLE she could develop many aspects of the Projective Reflection constructs in a way that made it similar to her starting point. This time her responses were more realistic in the sense that she gave less affirmative but more detailed responses to the prompts which were asked initially.

**Table 1: Summary of findings from the iteration 1 group and from an illustrative student case**

| <b>PR Construct</b>                 | <b>Starting self</b><br>(initial reflections; Weeks 1-2)   | <b>Possible selves explored</b><br>(during AVLE; week 3-7)  | <b>New self</b><br>(concluding reflections; week 8-9)  |
|-------------------------------------|--|---|--|
| Knowledge                           | Described experiences with digital tools, provided general, factually correct responses to knowledge questions   | Demonstrated increasingly detailed knowledge of urban planning (i.e. understanding relationships between specific design variables), and strong literacy using <i>PLS</i> compared to peers | Shared clear, detailed knowledge of urban planning concepts and general design process   |
| Interests and valuing               | Shared that she valued “making the environment we live in clean”   | Expressed enjoyment of AVLE experience, regularly identified relevance of urban planning to desired future career in science  | Described expanded interests in not only environmental conservation but also “the man-made items that surround it.”              |
| Self-organization and self-control  | Shared enjoyment of group work and confidence in her ability to complete learning tasks  | Participated as a group leader - volunteered design strategies in group meetings, offered technical assistance to peers, described feedback she received and her strategies for improvement | Affirmed how her participation “helped me see what it was like being an urban planner”   |
| Self-perception and self-definition | Hoped to work “in the science field” as an engineer or scientist; recognized that this future career means she would “have to [have] contributed to some type of urban planning” | Shifting towards a future career in medicine, stating “urban planning would affect me in a way because of jobs and how there would be enough places to go and study medicine.”              | Hoped to be “a doctor, scientist, astronomer;” strongly agreed when asked if she could see herself as an urban planner in future |

## Discussion

Case findings and group themes are summarized in Table 1. Synthesis of Andrea’s changes over time revealed *integrated* identity exploration through engagement with the urban planning process: the development of increasingly-specific content knowledge and AVLE literacy, repeated and diverse connections to personal interest and values and an enjoyment of the experience, active peer-to-peer leadership and the enactment of self-regulation strategies, and connections to her current and desired future roles as they emerged. As a result, Andrea was consistently able to find personal value and enjoyment in relation to her perceived self and active knowledge gaining over time. These connections demonstrate the potential of the augmented AVLE for promoting deeper learning integrated into learners’ understandings of their own interest and values, patterns of self-organization and self-control, and self-perception and self-definition. As such, the AVLE, which leveraged the affordances of both virtual environments and real-world augmentations (i.e. roleplay and discussion) has the capacity to promote a comprehensive identity exploration trajectory as defined by the four PR constructs.

Group findings suggest that the AVLE experience encouraged at least some aspects of identity exploration around urban planning careers for all students – each player concluded the experience able to justify whether or not they wished to become an urban planner based on aspects of their AVLE participation. However, a review of group findings revealed that not all students enacted every aspect of PR and that many students approached identity exploration in different ways (i.e. increasing confidence versus frustration over time). These results illustrate the need for further inquiry into patterns of identity exploration so that additional supports may be designed in AVLEs to better support PR across learners with different starting self-characteristics (i.e. interests, knowledge, etc.).

## Theoretical and educational significance

Projective Reflection (PR) serves as a theoretical lens for supporting student learning as a process of identity exploration that leads to change over time. The ability of adaptive collaboration to engage in (a) the intentional reflection on the self, (b) the process of exploring possible future selves related to new careers and roles in collaboration with stakeholders and other group members, and (c) the enactment targeted and intentional steps toward a desired new self, will serve as a particularly valuable skill for learners preparing for an evolving 21<sup>st</sup> century workforce. The development and enactment of *Virtual City Planning* offers useful insights for addressing how learning experiences may be optimally designed using a robust theoretical framework to support targeted identity exploration and change processes (DeVane, 2010), and how students' trajectories of identity exploration and change in the same learning experience may differ (Foster, 2014). As practical and theoretical understandings of identity exploration trajectories emerge through this research, the capacity of learning practitioners to design early targeted supports for learners based on their starting self characteristics increases.

## References

- Bagley, E. A., & Shaffer, D. W. (2015). Stop talking and type: comparing virtual and face-to-face mentoring in an epistemic game. *Journal of Computer Assisted Learning*, 31(6), 606-622.
- Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. S. (2016). Digital games, design, and learning: A systematic review and meta-analysis. *Review of educational research*, 86(1), 79-122.
- Cobb, P., Confrey, J., DiSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational researcher*, 32(1), 9-13.
- Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory into practice*, 39(3), 124-130.
- DeVane, B. M. (2010). *Toward sociocultural design tools for digital learning environments: Understanding identity in game-based learning communities*. University of Wisconsin--Madison.
- Ferdig, R. E. (2006). Assessing technologies for teaching and learning: understanding the importance of technological pedagogical content knowledge. *British Journal of Educational Technology*, 37, 749-760.
- Foster, A. (2014). CAREER: Projective reflection: Learning as identity exploration within games for science. *Drexel University: National Science Foundation*.
- Foster, A., & Shah, M. (2016). Knew me and new me: Facilitating student identity exploration and learning through game integration. *International Journal of Gaming and Computer-Mediated Simulations (IJGCMS)*, 8(3), 39-58.
- Foster, A., Shah, M., & Barany, A. (2017). Facilitating learning as identity change through game-based learning. In Y. Baek (Ed.), *Game-Based Learning: Theory, Strategies and Performance Outcomes*. New York, NY: Nova Publishers.
- Shah, M., Foster, A., Barany, A., Cellitti, J., Duka, M., Swiecki, Z., Evenstone, A., Kinley, H., Quigley, P., & Shaffer, D.W. (2018). Learning and Identity in Virtual Learning Environments: Iterative Design and Implementation of Philadelphia Land Science. Paper to be presented at the 1<sup>st</sup> Connected Learning Summit, MIT Media Lab, Cambridge, MA, August 1-3, 2018.
- Kamarainen, A. M., Metcalf, S., Grotzer, T., & Dede, C. (2015). Exploring ecosystems from the inside: how immersive multi-user virtual environments can support development of epistemologically grounded modeling practices in ecosystem science instruction. *Journal of Science Education and Technology*, 24(2-3), 148-167.
- Kaplan, A., Flum, H. (2012). Identity formation in educational settings: A critical focus for education in the 21<sup>st</sup> century. *Contemporary Educational Psychology*, 37(3), 171-175.
- Kaplan, A., Sinai, M., & Flum, H. (2014). Design-based interventions for promoting students' identity exploration within the school curriculum *Motivational interventions* (pp. 243-291): Emerald Group Publishing Limited.
- Krippendorff, K. (2004). *Content analysis: An introduction to its methodology*. Sage Publications.
- Maxwell, J. A. (2013). *Qualitative research design: An interactive approach* (Third ed.). Thousand Oaks, Calif: SAGE Publications.
- Milgram, P., Takemura, H., Utsumi, A., & Kishino, F. (1995). *Augmented reality: A class of displays on the reality-virtuality continuum*. Paper presented at the Telemanipulator and telepresence technologies.
- Shaffer, D. W. (2017). *Quantitative ethnography*: Cathcart Press.

## Acknowledgments

This material is based upon work supported by the National Science Foundation. The views expressed are those of the researchers.

# Collaborative Talk Across Two Pair-Programming Configurations

Zarifa Zakaria, Danielle Boulden, Jessica Vandenberg, Jennifer Tsan, Collin F. Lynch and Eric N. Wiebe  
zzakari@ncsu.edu, dmboulde@ncsu.edu, jvanden2@ncsu.edu, jtsan@ncsu.edu,  
cflynch@ncsu.edu and wiebe@ncsu.edu  
North Carolina State University

Kristy Elizabeth Boyer, University of Florida, keboyer@ufl.edu

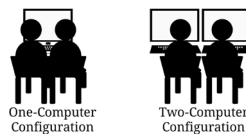
**Abstract:** Given that pair programming has proved to be an effective pedagogical approach for teaching programming skills, it is now important to explore alternative collaborative configurations. One popular configuration is where dyads collaborate by sharing a single computer sitting side-by-side. However, prior research points to potential challenges for elementary students when sharing a single computer when collaborating. This prompted us to explore another configuration where dyads sit side by side but collaborate on a shared virtual platform with individual computers. We compared the discourse of students' collaboration under these two settings. Results show that although there are no significant differences in the amount of collaborative talk between the two configurations, there is qualitative evidence of how differing affordances of two configurations shape collaborative elementary students' practices.

## Introduction

A plethora of studies suggest that collaboration is an effective pedagogical approach for teaching students programming skills (e.g., Hank et al. 2011; Williams et al., 2000) Additionally, the ability to collaborate effectively helps prepare students for the future workforce (National Research Council, 2013). However, there are many alternative configurations to implement collaborative programming in classrooms. For example, dyads can each use individual computers in a two-computer pair-programming configuration or a single computer to do one-computer pair programming (Figure 1). There can be differences in the capacity of collaboration under different physical configurations. In pair programming, a common configuration used in industry is sharing one computer (Hank et al. 2011). Here, one partner takes turns as a driver and the other as a navigator of the programming environment. Empirical studies exploring the benefits of such one-computer programming shows that students or employees working together on a single computer increases enjoyment (Muller, 2006), retention rate (Carver et al., 2007) and fast problem-solving (Xu & Rajlich, 2006).

While adapting industry practices for use in educational contexts makes sense, there is reason to believe that changes in context would need to be accounted for to optimize its benefits. First, studies have shown that collaboration can be improved in terms of enjoyment, engagement, and time-efficiency if there is equality of individual control over the computer (Infante, 2009; Scott, 2003). Secondly, the traditional one-computer configuration has mostly been studied either in industry or with college students (e.g., Salleh et al., 2011). For younger students there can be issues with equity (Shah et al., 2014; Tsan et al., 2018) which may hamper productive collaborative talk, raising questions as to the appropriateness of pedagogical strategies requiring a high degree of interpersonal negotiation like turn taking.

Thus, we aim to explore a two-computer configuration where partners collaborate in a virtual workspace sitting side-by-side at individual computers. There are studies with similar configurations: collaboration with multiple mice on a shared screen (Echeverría, 2012; Infante, 2009; Scott et al., 2003), multiple touch options on shared screens such as tablets (Hsiao et al., 2014), or individual screens (Cockburn, 2005; Dewan et al., 2009). However, little is known about collaboration with individual computers sharing a virtual space in close proximity (i.e., two-computer configuration). This configuration affords individual agency with system input but allows face-to-face dialogue with a partner. Our goal is to contrast the discourse level in such two-computer with one-computer collaboration in an emerging area of programming in K-12 academics.



**Figure 1.** The One- and Two-computer Configurations in Pair Programming.

## Theoretical framework

We focus our study on differentiating students' collaborative discourse throughout two different configurations of collaboration. When doing so, it is important to explore constructs that consider how the differences in these configurations might shape interactional levels of collective problem-solving. Mercer proposed the concept of Intermental Development Zone (IDZ) that focuses on the nature of interactive processes between teacher and student or peer-to-peer (Mercer, 2000). It draws on both Vygotsky's Zone of Proximal Development (ZPD) (Vygotsky, 1980) and Bruner's Scaffolding (Wood et al., 1976), to explain productive talk between peers. In a joint activity, three types of conversations can occur which Mercer termed 'Cumulative talk,' 'Disputational talk' and 'Exploratory talk.' In *Cumulative* talk, speakers build positively but uncritically on what the other has said, while *Disputational* talk is characterized by disagreement and individualized decision making. In *Exploratory* talk, participants engage critically, but productively, by challenging and offering alternative hypothesis to each other's ideas. Exploratory talk has been shown to expand the joint ZPD by enabling partners to achieve a better mutual understanding of the problem (Fernandez et al., 2015). Other research has shown that challenging viewpoints can foster conceptual change (Clark et al., 2003). As language is a prime medium of collaboration, even in computer-supported collaboration (Dillenbourg et al., 2009), it is appropriate to utilize Mercer's framework to explore language artifacts (i.e., spoken utterances) of dyadic collaboration.

Results from this study will explore how the two configurations might facilitate computational problem-solving during programming activities by quantitatively and qualitatively comparing these configurations. We sought to contrast the conversational attributes of these two configurations by documenting the three types of collaborative talk situated in the IDZ framework. We also describe a process of translating the framework into practice in the context of an elementary level collaborative programming environment.

Guided by Mercer's (2000) IDZ conceptual framework and prior empirical findings, we explore the following research questions: 1) What is the balance of different types of collaborative talk under different pair-programming configurations? 2) How are these types of talk different or similar across the two pair-programming configurations?

## Method

### Participants

This study is part of a design and development project investigating collaborative programming in elementary schools. Participants were from a fifth-grade classroom of a suburban school in the Southeastern United States. All the students were in an Academically and Intellectually Gifted program. We used data on 11 students (10-11 years), among whom 4 (36%) were girls and 7 (63%) were boys. In total, we had 7 dyads in one-computer and 6 dyads in two-computer configuration. Students were paired in alternative combinations. No set of two students were paired more than once. Except for one, all students participated in activities under both the configurations.

### Procedure

Students participated in problem-solving activities with a 5-day programming curriculum developed by the authors. In these activities, students used the NetsBlox block-based programming environment to develop a program by dragging blocks and combining them. All the pairs solved problems of the same difficulty level. Dyads collaborated using the two different configurations on alternating days. In the one-computer configuration, students were informed that one partner would act as the driver who would control the input devices, while the other partner, the navigator, would talk to the driver about the problems and anticipate actions. The teacher used a timer to make sure each partner had equal opportunity to be a driver and a navigator. Dyads in the two-computer configuration shared a virtual platform with individual computers.

### Data collection and analysis

The data source was video recordings of dyads collaborating on solving a programming problem. In the one-computer configuration, video recordings captured both the students' faces, voices, and the computer screen using the Open Broadcaster Software. For the two-computer configuration, the same software recorded each students' face, voice, and screens separately which were then merged using Adobe Premiere for analysis. Collaboration sessions lasted an average of 53 minutes. We utilized around 20 minutes of each dyad's collaboration. Criteria for selecting a section of videos were: audio-video clarity, comparable difficulty level of the activities, and if both the partners took turns as driver and navigator in a one-computer configuration.

We analyzed the data using a mixed method approach (Creswell & Creswell, 2017). First, we coded the videos to identify the distribution of three types of talk and analyzed these distributions using non-parametric



tests. After the quantitative phase, we employed qualitative analysis to identify patterns and complexities. For the qualitative analysis, the video-recorded dialogue of all pairs and their respective schematic coding were qualitatively analyzed as a multiple case study design (Yin, 2014). To guide these analyses, we relied upon our theoretical propositions using Mercer’s IDZ framework and prior literature on pair programming to aid in yielding descriptions of the contextual conditions of each thematic case. These descriptions enabled us to offer plausible explanations for the results from the quantitative phase.

### Categorization and coding

Our qualitative analysis was initially inspired by a study conducted by T’sas (2018) that focused on the three types of talk from IDZ framework to classify conversation. Guided by T’sas’ articulation of the talk types, a scheme aligned with Mercer’s original framing was utilized. Further piloting refined T’sas’ approach by changing a few problematic sub-categories. For example, ‘counter-challenge’ changed to ‘counter-challenge followed by consensus’ and placed in Exploratory talk, while ‘counter-proposition’ changed to ‘counter proposition with no consensus’ and put in Disputational talk. Also, utterances in a time interval which did not fall into either Exploratory or Disputational category were considered as Cumulative talk given that they were characterized without any disagreement or critical exchange and thus considered as assertive, yet uncritical. The finalized coding scheme consisted of three categories: Exploratory, Cumulative and Disputational (Table 1).

A professional transcription service transcribed the video recordings for coding. We utilized a time interval coding method (Bakeman, 2000) similar to that used in some other dialogue studies (e.g., Baines et al., 2009) to code not just an utterance but an aggregated series of responses in an interval, representing a particular type of talk. Piloting determined that 10-second intervals optimized capture of unique phrases with appropriate context. Thus, we divided a 20-minute recording into 120 ten second intervals and then labeled each interval exclusively to one category. Utilizing concurrent video with the transcripts allowed us to leverage facial expressions, body postures, and gestures to interpret the transcript. After training on the coding scheme, three authors coded 50% of the videos together (average Cohen’s *kappa* = 0.805) and then coded the other 50% individually. Coding discrepancies were then discussed until consensus was reached.

Table 1: Definitions of the Categories and *k*-agreements

|                            | Exploratory   | Cumulative   | Disputational  |
|----------------------------|---|--|--|
| Major Characteristics      | Challenge, alternative hypothesis, critical reasoning. <i>k</i> =0.714  | Uncritical addition of ideas, Agreement. <i>k</i> =0.893   | Disagreement without critical reasoning, <i>k</i> =0.803   |
| Elaborated Characteristics | <ul style="list-style-type: none"> <li>- Offered alternative hypothesis.</li> <li>- Initiations challenged, and counter-challenged followed by consensus.</li> <li>- Justifications given.</li> <li>- Joint acceptance</li> </ul> | <ul style="list-style-type: none"> <li>-Agreement without critical discussion.</li> <li>-Friendly and conflicts avoidance.</li> <li>-Positively but uncritically sharing ideas.</li> <li>-Superficial amendments.</li> </ul> | <ul style="list-style-type: none"> <li>-Disagreement without outcome.</li> <li>-Individualized decision-making.</li> <li>-Initiations directly rejected.</li> <li>-No/little constructive criticism.</li> <li>-Counter proposition with no consensus.</li> <li>-No resolutions.</li> </ul> |

### Results and discussion

To answer our first research question, we examined the distribution of talk by quantitatively examining the coded frequencies of the three types of talk. We first investigated the overall level of discourse between the two configurations. The unit of overall quantitative analysis was students’ talk per minute. We took each coded talk in 10 second intervals and aggregated them in number of different types of talk per minute. Non-parametric statistical test results indicated a non-significant difference,  $U=20.5$ ,  $p=.943$ , between the two-computer configuration (66.9%) and one-computer (81.7%). The trend of a lower percentage of talk in the two-computer configuration could be explained by the fact that each student had equal autonomy in terms of computer access, limiting the need for negotiation. A Kruskal-Wallis test was used in an analysis into the distribution across the percentage of different types of talk per minute, and showed a statistically significant difference,  $H= 24.9$ ,  $p<.001$ . Distribution was 12.7% Exploratory, 82.6% Cumulative and 4.7% Disputational. A series of Wilcoxon tests between pairs of each of the talk types revealed the following: Exploratory-Cumulative,  $Z=-3.18$ ,  $p<.01$ , Cumulative-Disputational,  $Z=-3.18$ ,  $p<.01$ , Exploratory-Disputational,  $Z=-2.85$ ,  $p<.01$ . Previous literature

indicates that, generally, without any intervention, Cumulative and Disputational talk tend to be higher than Exploratory (T'sas, 2018). While no significant differences were found between the configurations for any of the types of talk, Disputational trended higher for one-computer (7.0%) than two-computer (2.6%) (Figure 2).

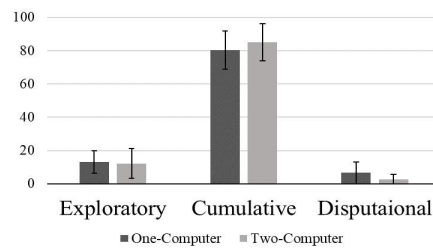


Figure 2. Percentage of each type of talk per minute, by computer configuration.

Our second research question allowed us to explore the quality of the talk across the conditions. We qualitatively analyzed the collaborative conversations and found some salient aspects of talk which elucidate differences between collaboration in one- and two-computer configurations. From the coded talk, we found a few sub-categories like, “challenge”, “explain”, “cooperation’ etc. Three themes emerged from these categories.

### Exploratory talk often preceded/followed by Disputational talk in One-computer

Results of qualitative analyses suggest that the transition between Exploratory talk and the other types of talk differed based on the configural conditions. We provide examples to better illustrate this phenomenon.

Tom and Sandy (pseudonyms) present one example of a one-computer configuration where Exploratory talk was preceded by Disputational talk. They repeatedly disagree with one another without any explanation to each other. There are disagreements like "No" and "Nope" multiple times without trying to come to a consensus. This falls into the characteristic of Disputational talk. However, after several statements of dispute, Tom hypothesizes, “that is as far as it’s gonna go,” whereby Sandy challenges, “How do you know?” This challenge forces him to counter her question with further explanation which becomes Exploratory talk.

| One-Computer   | Two-Computer   |
|--|--|
| <p><u>Tom</u>: Nope, not far enough, negative 150.<br/> <u>Sandy</u>: Negative 110.<br/> <u>Tom</u>: No, negative 150.<br/> <u>Sandy</u>: Let's go with 10, maybe. 10?<br/> <u>Tom</u>: Nope, five.<br/> <u>Sandy</u>: One. So we need to go, like, negative-<br/> <u>Tom</u>: Well, no, it's- it's going to X.215.<br/> <u>Sandy</u>: Negative ...<br/> <u>Tom</u>: That's as far as it's gonna go.<br/> <u>Sandy</u>: How do you know?<br/> <u>Tom</u>: Because it's a negative. You see, 'cause that's still the same number.</p> | <p><u>Tony</u>: What'd you do?<br/> <u>Sandy</u>: I made, whoa, it's supposed to go right when I hit the right arrow but instead, it's like flying away.<br/> <u>Tony</u>: Yeah.<br/> <u>Sandy</u>: Why though?<br/> <u>Tony</u>: Oh, that happens.<br/> <u>Sandy</u>: Why?<br/> <u>Tony</u>: I don't know if it's just 'cause we don't understand what's happening. Because whenever we did that, it'd change that too much and it'd make it. I forgot how we did it last time.<br/> <u>Sandy</u>: It might have something to do with that.</p> |

Figure 3. Excerpts illustrating the relationship of Disputational talk with Exploratory (one-computer) or Cumulative (two-computer).

In comparison, consider these transitions between types of talk in two-computer configurations, where Cumulative talk progressed into Exploratory talk and then is followed by more Cumulative talk. In the excerpt in Figure 3, the same girl, Sandy, is paired with another boy, Tony. Here, Exploratory talk originates from Cumulative talk where Sandy requests Tony’s assistance by saying “Tony, help me.” This request then segues into a repetitive pattern of Exploratory phrases comprising a series of stimulating questions when each student responds to their partner’s inquiry using several ‘why’ questions and ‘because’ explanations.

These examples are demonstrative of a divergent sequential form for these two conditions, where in the one-computer configuration Exploratory talk transitioned into Disputational talk while in the two-computer configuration Exploratory talk transitioned back into Cumulative talk. A possible explanation may be that because students in two-computer have equal access to their own devices, allowing them agency to express their ideas directly without requiring negotiation, the likelihood of control-based dispute is less likely to ensue.

Past research on student conflict during collaborative activities has demonstrated that this is a naturally occurring phenomenon (Jeong, 2008; Rubin, Pruitt, & Kim, 1994), and can be beneficial or harmful to outcomes

depending on its characteristics. For example, Lee, Huh, and Reigeigluth (2015) found task-related conflict to have a positive influence on peer collaboration as students are forced to self-reflect and build upon one another's ideas. Whereas they contend that when students are having disputes about the collaboration process itself or interpersonal conflicts, it can be detrimental to productive collaboration. There is concern that the one-computer configuration is perhaps amplifying the potential for unproductive disputes over process or based on interpersonal dynamics. The two-computer configuration still requires negotiation over the final programming artifact but removes the need for sharing input devices, perhaps striking a better balance of agency and control.

### Balance of challenger and explainer in Exploratory talk

In the one-computer configuration, we found evidence that one learner may end up challenging their partner more by asking questions and demanding justification for any proposals or claims; we consider this type of learner to be a *challenger*. In the excerpts below, we can see the other partner took the role of an *explainer* and ended up justifying what she proposed or edited. As challenging is identified as a major characteristic of Exploratory talk, this pairing of *challenger* and *explainer* was found in this type of talk. Interestingly, among the excerpts, where this combination was found, it was the driver (who had control of the input) who played the role of challenger. With Steve driving and Melony navigating, Steve was more of a challenger, asking more questions to Melony who was answering them. In one instance, they were discussing deleting a block and Steve (driver) was constantly challenging Melony's idea with questions such as "Okay. So, what?", "How do you have two layers?", "Like, does it still work?", "Okay, So what?" one after another (see Figure 4). This tendency of the driver to be more challenging was found in other dyads in the one-computer configuration too. In another pair, Sandy, the driver, was challenging Anthony's ideas with a lot of 'why' and 'how' questions. Previous research suggests that 'why' and 'how' are Exploratory key words for being counter challenging and 'because', 'but', 'what if' are Exploratory keywords for explanation (Polo et al., 2015; Wegerif et al., 2005).

| One-Computer  | Two-Computer  |
|---|---|
| <p><u>Steve</u>: Okay. So, what?<br/> <u>Melony</u>: And so, like here, wouldn't you do, um..<br/> <u>Steve</u>: How do you have two layers?<br/> <u>Melony</u>: I don't know.<br/> <u>Melony</u>: Maybe you've to put wind direction off.<br/> <u>Steve</u>: Like, does it still work?<br/> <u>Melony</u>: Yeah, it works, but like ...<br/> <u>Steve</u>: Okay. So, what?<br/> <u>Melony</u>: And so, like here, wouldn't you do,um..</p> | <p><u>Tony</u>: Because it was like this, when we pressed up arrow ... oh wait he has a...<br/> <u>Steve</u>: But I think we're actually connected these, when you press it this goes to-<br/> <u>Tony</u>: What happen if would go down more than he would go up?<br/> <u>Steve</u>: He doesn't even go down though.</p> |

Figure 4. Excerpts illustrating the balance of challenger and explainer in Exploratory talk.

In contrast, in the two-computer configuration, there was evidence that showed both the partners taking the roles of challenger and explainer in a more interleaved fashion. In Figure 4, Tony and Steve are found to be continuously challenging and explaining to each other with words such as 'but,' 'because,' 'what happens if,' 'even though' etc. Another pair, Luke and Clara were similarly questioning each other's ideas and demanding an explanation. In one instance, Clara was suggesting an 'if/else' block as well as challenging Luke's idea of inputting a 'stop' block. Luke too, was simultaneously proposing the opposite idea and kept questioning Clara.

This differing combination of *challenger* and *explainer* suggests that this relationship is not always equitable with the challenger playing more of an authoritative role in the pair. Researchers have strived to analyze and understand the equity of collaborative relationships in many subjects, including computer science (e.g., Deitrick et al., 2016; Shah et al., 2014; Stanton et al., 2002). While we are still unsure of the effects of an inequitable relationship in terms of the challengers and explainers, we hypothesize that collaboration is productive if the dyads have balanced, interleaved challenging and explaining interactions.

### Cooperative work rather than collaborative work in two-computer configuration

In the two-computer configuration, both partners have access to a laptop, thus they get the opportunity to edit individually. This increase possibilities of getting involved in *cooperative* work rather than *collaborative* work, which has a clear distinction in terms of the socio-cultural learning mechanism (Davidson, 2014; Oxford, 1997). Kagan and Kagan (2009) identify a pair of critical attributes of cooperative learning: interdependence and accountability. Here, students can work in separate sections of the project, yet still simultaneously seek help from each other. For example, Rupert and Melony divided the work between themselves by saying "Yeah, we need to.

I'll do the top one, you can do the bottom one," indicating different code blocks. However, they both sought help when they came upon problems. Melony, facing difficulty, asked "In this one? And then the other way around?". Rupert then assisted her, answering "And then you can put negative 85 in the second one." This pattern of interdependence then continued. Figure 5 shows another prominent example, where Sandy and Tony start cooperating, Sandy responds to a teacher saying, "I'm gonna work on the sprite getting it to move back and forth and he's gonna work on the different enemies." However, after a few minutes of cooperative work, they start going back and forth to each other's section, hovering over their partner's computer to help them edit their sections. While doing cooperation, they were more involved in Cumulative talk. But when they started collaboration on same section of the code, more Exploratory instances were observed.

| Two-Computer (Cooperative)   | Two-Computer (Collaborative)   |
|--|--|
| <p><u>Sandy</u>: You work on the ball, I'll work on the platforms.<br/> <u>Tony</u>: Let me just duplicate this last group and then add it.<br/> <u>Sandy</u>: Wait, I'll work on the ball actually, you keep going the enemies.<br/> <u>Tony</u>: Okay, let me fix. okay, so, he shouldn't be doing this forever now. Let me check.<br/> <u>Tony</u>: Yay, I've got enemy one. But now I have enemy two doing the exact same thing.</p> | <p><u>Sandy</u>: Broadcast. We're gonna need some sort of broadcast in there.<br/> <u>Melony</u>: Okay, broadcast...Scissors...<br/> <u>Sandy</u>: Oof! Hello. Oof! Why is it repeating that? Oh, no, what did we do? We broke it! We broke it. We broke it.<br/> <u>Melony</u>: Don't worry, I know why.<br/> <u>Sandy</u>: Is it flashing for you, too?<br/> <u>Melony</u>: Yeah. I put it in forever. There. Oh, no. There we go. Now it'll only say it once. See?<br/> <u>Sandy</u>: But then it just disappears. What did you do, Melony?<br/> <u>Melony</u>: What?</p> |

Figure 5. Excerpts illustrating the individual vs. collaborative talk in two-computer configuration.

In the two-computer configuration, we also found both the students working on the same sprites but different blocks. The collaborative excerpt in Figure 5 shows Sandy and Melony working on the same section of code, challenging and helping each other as they work through it. While, as we see above that the two-computer collaboration could devolve into cooperative work, dyads seldom did work completely by themselves, rather they often verbalize what they were doing and sought help. Thus, these interactions are best characterized as moving along a collaborative-cooperative continuum. Relevant literature supports the idea that once each student has their own input device, they tend to cooperate rather than collaborate (Stanton et al., 2002).

Cooperative learning has been found to be an effective learning strategy to enhance achievement (McConnell, 2014) and thus a reason for exploration in pair-programming configurations. For future research we would explore productive outcomes of pair-programming with cooperative or collaborative interactions.

## Conclusion

Guided by research informed by Mercer's (2000) framework, we believe helping students to productively collaborate with a higher percentage of Exploratory talk is a laudable goal when designing a pair-programming platform for elementary students. Exploratory talk offers potential for learning where evidence was found to extend students' learning within the ZPD (Littleton, 2005). Our findings that Exploratory talk occurred much less frequently than Cumulative were congruent with past studies (Bungnum, 2018; T'sas, 2018), which showed that without any intervention most conversation generally fall into positive but uncritical utterances, characteristic of Cumulative talk. However, our focus will be to use a combination of system and teacher level supports that enhance the quality and quantity of productive, Exploratory talk.

While there were no significant quantitative differences in the three types of talk between the configurations, qualitative analysis revealed interesting patterns that possibly point to how one and two-computer configurations shape collaborative work by students of this age. The one-computer configuration that provided direct control of editing code to only one student seemed to create an imbalance in an agency with regards to both proposing and evaluating potential solution paths. Evidence pointed to challenging ideas residing primarily with the driver, with the navigator left defending their ideas. We wonder if this is related to the pattern of Disputational talk proceeding Exploratory talk more often in the one-computer configuration. This result was not surprising because the driver is the learner that has control over the computer, thus control over the changes in the code. While we believe that the inequity of the challenger-explainer interaction may be due to the nature of the pair-programming roles, we plan to conduct statistical comparison to verify the differences between the two configurations. Our goal will be to develop interventions and adaptive support to improve the equity of these types of interactions. Examples of interventions include collaboration trainings to educate students on how to challenge their partner's idea and having intelligent collaboration agents that can urge both the students to get involved in challenging and explaining. Of course, one way to limit conflict over editing the program is to provide both students with input control. However, that configuration has the risk of dissipating the collaborative relationship,

as evidence suggested that collaboration in two-computer configuration has a possibility of turning into cooperation. That said, our findings point to a healthy cooperative interdependence of the students in the two-computer configuration, as evidenced by the interleaving of challenger-explainer roles and regular consultation when working on separate codes.

In future research, we will continue to examine and operationalize Mercer's framework for measuring and analyzing dialogue in pair programming. We want to explore Cumulative talk more exhaustively to identify if there are methods of disaggregation of the Cumulative category that are parsimonious with theory and useful empirically to our research and development. Larger scale data collection with more students will also allow us to reach more conclusions about the quantity of both Exploratory and Disputational talk between the two conditions. These findings also have important implications for our future design efforts on systems for pair programming where virtual agents could be used to facilitate the use of Exploratory talk.

## Limitations

Our study consisted of a relatively small, non-representative sample of students, limiting our statistical power and ability to generalize findings. Second, we were unable to analyze partner differences by condition due to instructional and logistical limitations. Relatedly, our analysis did not attempt to account for individual differences (e.g., prior programming knowledge, gender, etc.) which might have impacted dyad's collaborative dynamics. Future studies are planned to address these limitations.

## References

- Bakeman, R. (2000). Behavioral observation and coding. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 138-159). New York, NY: Cambridge University Press.
- Baines, E., Rubie-Davies, C., & Blatchford, P. (2009). Improving pupil group work interaction and dialogue in primary classrooms: results from a year-long intervention study. *Cambridge Journal of Education*, 39(1), 95-117.
- Bungum, B., Bøe, M. V., & Henriksen, E. K. (2018). Quantum talk: How small-group discussions may enhance students' understanding in quantum physics. *Science Education*, 102(4), 856-877.
- Carver, J. C., Henderson, L., He, L., Hodges, J., & Reese, D. (2007, July). Increased retention of early computer science and software engineering students using pair programming. In *Software Engineering Education & Training, 2007. CSEET'07.* (pp. 115-122). IEEE.
- Clark, A. M., Anderson, R. C., Kuo, L. J., Kim, I. H., Archodidou, A., & Nguyen-Jahiel, K. (2003). Collaborative reasoning: Expanding ways for children to talk and think in school. *Educational Psychology Review*, 15(2), 181-198.
- Cockburn, A. (2005). Crystal Clear. Addison-Wesley.
- Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Thousand Oaks, CA: Sage.
- Davidson, N., & Major, C. H. (2014). Boundary crossings: Cooperative learning, collaborative learning, and problem-based learning. *Journal on Excellence in College Teaching*, 25.
- Deitrick, E., Shapiro, R. B., & Gravel, B. (2016). How Do We Assess Equity in Programming Pairs? In Looi, C. K., Polman, J. L., Cress, U., and Reimann, P. (Eds.). *Transforming Learning, Empowering Learners: The International Conference of the Learning Sciences (ICLS) 2016*, Volume 1. Singapore: International Society of the Learning Sciences..
- Dewan, P., Agarwal, P., Shroff, G., & Hegde, R. (2009, May). Distributed side-by-side programming. In *Proceedings of the 2009 ICSE workshop on cooperative and human aspects on software engineering* (pp. 48-55). IEEE Computer Society.
- Dillenbourg P., Järvelä S., Fischer F. (2009) The Evolution of Research on Computer-Supported Collaborative Learning. In: Balacheff N., Ludvigsen S., de Jong T., Lazonder A., Barnes S. (eds) *Technology-Enhanced Learning*. Springer, Dordrecht.
- Echeverría, A., Améstica, M., Gil, F., Nussbaum, M., Barrios, E., & Leclerc, S. (2012). Exploring different technological platforms for supporting co-located collaborative games in the classroom. *Computers in Human Behavior*, 28(4), 1170-1177.
- Fernández, M., Wegerif, R., Mercer, N., & Rojas-Drummond, S. (2001). Re-conceptualizing "Scaffolding" and the Zone of Proximal Development in the Context of Symmetrical Collaborative Learning. *The Journal of Classroom Interaction*, 36/37(2/1), 40-54.
- Hanks, B., Fitzgerald, S., McCauley, R., Murphy, L., & Zander, C. (2011). Pair programming in education: a literature review. *Computer Science Education*, 21(2), 135-173.

- Hsiao, H. S., Chang, C. S., Lin, C. Y., Chang, C. C., & Chen, J. C. (2014). The influence of collaborative learning games within different devices on student's learning performance and behaviours. *Australasian Journal of Educational Technology*, 30(6).
- Infante, C., Hidalgo, P., Nussbaum, M., Alarcón, R., & Gottlieb, A. (2009). Multiple Mice based collaborative one-to-one learning. *Computers & Education*, 53(2), 393-401.
- Jeong, H. W. (2008). *Understanding conflict and conflict analysis*. Thousand Oaks, CA: Sage.
- Kagan, S. (2009). *Kagan's cooperative learning*. San Clemente, CA: Kagan Publishing.
- Mercer, N. (2000). *Words and minds: How we use language to think together*. New York, NY: Routledge.
- Müller, M. M. (2006). A preliminary study on the impact of a pair design phase on pair programming and solo programming. *Information and Software Technology*, 48(5), 335-344.
- Lee, D., Huh, Y., & Reigeluth, C. M. (2015). Collaboration, intragroup conflict, and social skills in project-based learning. *Instructional Science*, 43(5), 561-590.
- Littleton, K., Mercer, N., Dawes, L., Wegerif, R., Rowe, D., & Sams, C. (2005). Talking and thinking together at Key Stage 1. *Early Years*, 25(2), 167-182.
- Oxford, R. L. (1997). Cooperative learning, collaborative learning, and interaction: Three communicative strands in the language classroom. *The Modern Language Journal*, 81(4), 443-456.
- McConnell, D. (2014). *Implementing computing supported cooperative learning*. Routledge.
- National Research Council. (2013). *Education for life and work: Developing transferable knowledge and skills in the 21st century*. Washington, DC: National Academies Press.
- Polo, C., Lund, K., Plantin, C., & Niccolai, G. (2015). *Analyzing Exploratory Talk as a Socio-Cognitive Practice: Identity, Group Argumentation, and Class Debate Quality*. (hal-01208319)
- Rubin, J. Z., Pruitt, D. G., & Kim, S. H. (1994). *Social conflict: Escalation, stalemate, and settlement*. New York, NY: McGraw-Hill.
- Salleh, N., Mendes, E., & Grundy, J. (2011). Empirical studies of pair programming for CS/SE teaching in higher education: A systematic literature review. *IEEE Transactions on Software Engineering*, 37(4), 509-525.
- Scott, S. D., Mandryk, R. L., & Inkpen, K. M. (2003). Understanding children's collaborative interactions in shared environments. *Journal of Computer Assisted Learning*, 19(2), 220-228.
- Shah, N., Lewis, C., & Caires, R. (2014). Analyzing equity in collaborative learning situations: A comparative case study in elementary computer science. In *Proceedings of the International Conference of the Learning Sciences*, 495-502. International Society of the Learning Sciences. Boulder, CO.
- Stanton, D., Neale, H., & Bayon, V. (2002, January). Interfaces to support children's co-present collaboration: multiple mice and tangible technologies. In *Proceedings of the Conference on Computer Support for Collaborative Learning: Foundations for a CSCL Community*, 342-351. International Society of the Learning Sciences.
- Tsan, J., Lynch, C. F., & Boyer, K. E. (2018). "Alright, what do we need?": A study of young coders' collaborative dialogue. *International Journal of Child-Computer Interaction*, 17, 61-71, ISSN 2212-8689.
- T'Sas, J. (2018). *Learning outcomes of exploratory talk in collaborative activities* (Doctoral dissertation, University of Antwerp).
- Vygotsky, L. S. (1980). *Mind in society: The development of higher psychological processes*. Harvard University Press. Cambridge, Massachusetts.
- Wegerif, R., Linares, J., Rojas-Drummond, S., Mercer, N., & Velez, M. (2005). Thinking Together in the UK and Mexico: Transfer of an Educational Innovation. *The Journal of Classroom Interaction*, 40(1), 40-48.
- Williams, L., Kessler, R. R., Cunningham, W., & Jeffries, R. (2000). Strengthening the case for pair programming. *IEEE Software*, 17(4), 19-25.
- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89-100.
- Xu, S., & Rajlich, V. (2006, July). Empirical validation of test-driven pair programming in game development. In *Computer and Information Science, 2006 and 2006 1st IEEE/ACIS International Workshop on Component-Based Software Engineering* (pp. 500-505). IEEE.
- Yin, R. K. (2017). *Case study research and applications: Design and methods*. Thousand Oaks, CA: Sage.

## Acknowledgements

This work is supported in part by the National Science Foundation (NSF) through the grant DRL-1721000. Everything expressed in this material are those of the authors and do not necessarily reflect the views of NSF.

# CoProStory: A Tangible Programming Tool for Children's Collaboration

Xiaozhou Deng, Danli Wang, and Qiao Jin

dengxiaozhou2018@ia.ac.cn, danli.wang@ia.ac.cn, jinqiao2018@ia.ac.cn

The State Key Laboratory of Management and Control for Complex Systems,

Institute of Automation, Chinese Academy of Sciences, Beijing, China

School of Computer and Control

Engineering, University of Chinese Academy of Sciences, Beijing, China

**ABSTRACT:** In this paper, we present CoProStory, a collaborative tangible programming tool to support children learning programming. In recent years, many collaborative tangible programming tools were developed for children. Compared to the other tools, Sync Blocks which were designed to coordinate children's programming were used in our tool. The tool consists of two parts: tangible programming blocks and game tasks. Game tasks are provided with 3D cartoon story scenes on computer screen and need to be finished by collaborative programming with our programming blocks. To evaluate our tool, a contrast user study was conducted with 28 children aged 6 to 10. The results demonstrated that CoProStory could support children to program collaboratively and children had a more positive attitude toward programming with Sync Blocks.

## Introduction

Children programming has been a wide field of research since 1960s. Research has shown that programming education can have a positive and measurable effect on children's achievement, not only in areas such as math and science, but also in language skills, creativity, and social emotional interaction (Douglas, 1999). Besides, learning programming is an efficient way to cultivate computational thinking which has been described as a fundamental skill for everyone, not only for computer scientists (Jeannette, 2006).

However, programming appears to be quite a challenge for young children. With the traditional way of programming, children have difficulties not only in learning rigid syntax and text symbols, but also in using the complex programming environment (Andy & Andrew, 2007). Therefore, lowering the difficulty of programming and offering children proper programming tools are highly valuable in education. TUI is an efficient interaction method embracing the richness of human's interaction with the physical world, which contributes to children's learning (Michael & Robert, 2007). With tangible programming tools, children can write programs by assembling the physical objects without keystrokes, which is much easier to involve children in programming (Michael & R.Jordan et.al, 2012; Timothy, 2004). Besides, TUI has the advantages of involving multiple children in the same process, as TUI can easily provide an open and shared environment among children (Hiroshi & Brygg, 1997).

Face-to-face collaboration with classmates or friends is an important part of children's daily lives. Collaboration provides children with an efficient way to improve social communication skills. It helps children analyze information from the perspectives of other users when communicating face-to-face (Regan & Kori et.al, 2001). Collaborative programming can promote the awareness of the teamwork, and the skills of communicating and collaborating (Scott, Regan & Kori, 2003). However, for children, if the tasks are not assigned well, the inequality of task loads can result in significant decrease of children's practice time, interests and concentrations. And, research has indicated that tangible user interfaces could provide broad and shared interaction environments, which have the potential to involve multiple children at the same time and support face-to-face collaboration (Leslie, 2007). Furthermore, research implicated that collaborative technology should support concurrent interaction which can help to engage children in a collaborative activity and enable them to participate equally (Neha & Laurie et.al, 2004; Charlie & Linda et.al, 2002).

Based on the analysis above, we applied the concept of process synchronization and developed CoProStory, a tangible programming system supporting children's collaboration in programming (Figure 1). Compared with previous works, CoProStory allows two children to program their own characters with programming blocks in parallel, which aims to complete a common task. In this method, with clear role division, conflicts could be well reduced. Furthermore, Sync Blocks are provided to coordinate children's collaboration.



**Figure 1.** Children playing with CoProStory.

## Related work

One of the earliest tangible programming projects that support collaboration is AlgoBlock (Hideyuki & Hiroshi, 1995). Children write their own programs to play a marine game by connecting the objects. In their user study, the authors proved AlgoBlock's value in improving the communication among children by recording and analyzing their speech and activities. Tangicons 3.0 is an educational game designed for children age 6 - 9 (Florian & Thomas et.al, 2012). It provides a collaboration environment by allowing at most four children to program together to move virtual characters on a map. However, its programming blocks are built based on Sifteo cubes, which increase the cost of the tool. Digital Dream Lab (DDL) is a tangible storytelling system for children to learn programming concepts. DDL focused on the concepts of variables and classes by using the metaphor provided by irregular puzzle pieces (Hyunjoo & Anisha, 2013). The system has six kinds of blocks, including Character Block, Animation Block, Color Block, Size Block, Variable Block and Background Block. However, the tool was also designed for older children (7-12 years old) and younger children are short in collaboration compared to those of older children. Plugramming was designed to support young children's face-to-face collaborative programming. However, in this tool tasks are divided into two separated branches of a program (Tomohito & Yasushi et.al, 2017).

From the works above, we find the open characteristics of TUIs can effectively support face-to-face collaboration between children. Based on previous work, we designed and developed a new tangible programming system, CoProStory, which allows two children play and program face-to-face. With clear role division and Sync Blocks, CoProStory could support two children to program concurrently and collaboratively.

## Design and implement

Our design and implement process involved two major revisions in which we tried to explore how to encourage children's collaboration better. The difference between the two versions is that one is designed to be completed with Sync Blocks while the other isn't. The details are described as follows:

### Programming blocks

Programming blocks are small square boxes (6cm\*6cm\*6cm) covered with color figures representing certain semantics or manipulation constraints (Figure 2). Programming blocks are the primary tools for children interacting with the animation game running on the computer.

Blocks with RFID reader are very flexible, and are suitable for parameters with a lot of optional values. In these blocks, a RFID reader was fixed beneath the top surface. And with a shallow groove on its top surface, a mini Interactive Element Card (IECard) could be embedded, together making up a complete semantic. IECards are made up with small circular RFID tags covered with figures, which could represent different interactive elements in game scenes, such as a tree, a gate and even a character. And with infrared sensor fixed inside, the message of related position could be provided.

Programming blocks can be divided into four categories: Character Blocks, Attribute Blocks, Animation Blocks and Sync Blocks. Children could use Character Blocks and Attribute Blocks to initialize their characters, which provides with clear role division. Since Attribute Blocks provide children with individual changes to visual character, it may be easier to make children immerse themselves to our game. Animation Blocks are used



as commands to program the characters' actions. With different kinds of IECards, children could use Animation Blocks to program the visual character finishing the given tasks. Most importantly, we design a special type of blocks, Sync Blocks, to coordinate children's collaboration.

Sync Blocks are designed to encourage children's collaboration in the game. Three kinds of Sync Blocks are currently provided: Wait Blocks, Notice Blocks and WaitForSeconds Blocks. Wait Blocks and Notice Blocks are always used in our programs. Once it comes to a wait command, the corresponding character would stand there and keep waiting until another character comes to a notice command. To use the Sync Blocks properly, children need to decompose the given task together before programming and figure out a solution with provided blocks.

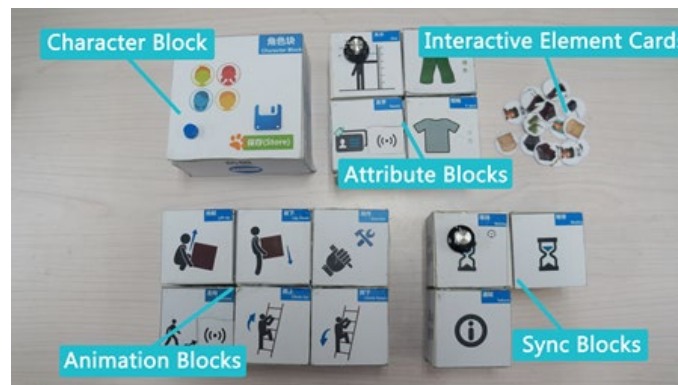


Figure 2. Programming Blocks & Interactive Element Cards.

Sync Blocks help children to keep in the same speed to some extent. If children ignore their partner's process, the real-time feedback would provide corresponding error messages. For instant, one character needs to cross a river but he should wait for the other to drop the suspension bridge and children should use Sync Blocks to complete these moves.

## Visual games

We provide two versions of games which need different degrees of collaboration. Specifically, the work in the first version is designed without Sync Blocks.

In the game of version1, children should use programming blocks to move their characters to the given destination. One of characters needs to go to the tree from the house while the others direction is opposite. During this process they should make their characters finish certain movements such as opening an umbrella or striding over a rock in order. They need to program the correct sequences without Sync Blocks.

In the game of version2, children need to finish their task with Sync Blocks. After several movements, one character should wait for the other dropping down the suspension bridge by a series of operations to cross the river. After the character got cross the river, the other character should lift the suspension bridge again. In this story scene, two characters' movements are time-related, as a result, children need to use Sync Blocks to finish this given task.



Figure 3. Different versions of game tasks.

## Programming process

The whole process of learning consists of three stages: initialization stage, programming stage and running stage. Children would use Character Blocks and Attribute Blocks to initialize their characters. Then simple background stories would be told to children which contains game task. After children get clear with their goals, they are asked to program characters to complete given tasks. During this programming stage, the real time feedback showed in figure 4 would help children and reduce their cognitive load of programming. Specifically, there would be states messages on the screen which could show children their progress in real time. And if children place wrong blocks, there would be error messages to help children debug their program. After children finish their program, the running state would show the result of children's programming with animation. We use these states to introduce children the normal process of programming.

To evaluate our tool, we conducted a contrast user study and compare these two versions to find the benefits of Sync Blocks.



Figure 4. Programming state and running state.

## User study

### Participants and setting

28 children (13 girls and 15 boys) aged 6 to 10 with a mean age of 7.79 were involved in our user study. To assess the collaborative design of CoProStory, 28 children were divided into two groups evenly: G1 (7 girls and 7 boys with mean age of 7.5) and G2 (6 girls and 8 boys with mean age of 8.14). Children in G1 were experimented with Version1 of CoProStory and children in G2 experimented with Version2. Experiments were conducted in a spacious room. In order to capture children's behaviors and voice, a video recorder was set up.

### Experimental procedure

The experimental procedure mainly consists of three parts: introduction, programming and interview. At the beginning, we introduced CoProStory to children (5 minutes). After that, children were asked to complete the programming tasks in pairs (30 minutes) and the interactions, conversations, utterances and time they spent on each task would be recorded with cameras. Next, we made a brief interview for children (10 minutes). Altogether, one experiment takes one regular school lesson (45 minutes).

### Measures

In order to find the advantages of the design of Sync Blocks with quantitative data, we conducted quantitative coding and analysis of video data of the participants' behaviors act by act using Bale's Interaction Process Analysis coding scheme (Robert, 1950). And in order to raise measurement accuracy, each video data would be analyzed by two coders. Firstly, the two coders would learn the coding schema and coding rules, then coded 2 groups of video data independently. Afterward, they discussed their disagreements, adjusted their mismatched codes and then coded the remaining data separately.

According to Bale's Interaction Process Analysis coding scheme, we used these four kinds of codes to evaluate the interaction of children while programming: 1) positive reactions, 2) negative reactions, 3) initiative helping behaviors, 4) behaviors of asking for assistance. With the quantitative data of these codes, we could measure the collaboration of children and compare the results of the two groups to find the positive effects of Sync Blocks. Table1 is the description of each code.

**Table 1: Coding scheme used for our analysis of the recorded video**

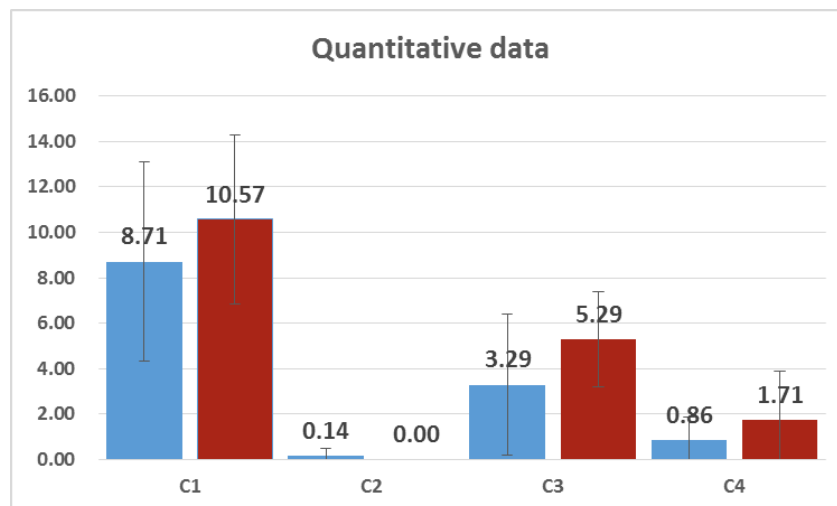
| Code | Description  |
|------|--|
| C 1  | Positive Reaction, including 4 kinds of behaviors: task-related conversation, task-related respond, passing blocks and correct partner's errors. |
| C 2  | Negative Reactions including refuse to help/answer and ignore partner's task-related talk and request.   |
| C3   | Take the initiative to provide help and answer task-related questions.   |
| C4   | Ask for help and ask task-related questions.   |

And we made a brief interview after children programmed. The interview questionnaire is a Likert-type scale composed of 8 questions scored from one to five, where one means the minimum and five means the max score. Q1-Q4 concerned about CoProStory's interests, ease to use and learn. Q5-Q8 are used to evaluate how CoProStory support children's collaboration. Questions are as follows:

- Q1: How much do you like the game?
- Q2: Do you think the game is easy to learn?
- Q3: Do you think it is easy to control roles in the game by blocks?
- Q4: Was the real-time feedback in the game helpful in the programming stage?
- Q5: How much do you like playing the game with your partner?
- Q6: Do you pay attention to your partner when you were playing the game?
- Q7: How much help did you offer your partner?
- Q8: How much help did your partner offer you?

## Results

We analyzed the quantitative data of four codes combined with video data. The quantitative coding results are shown in Figure 5:



**Figure 5.** Quantitative coding results of G1 (blue) and G2 (red).

For interactive behaviors, the average positive data (C1) of G1 and G2 are 8.71(SD=4.40) and 10.57(SD=3.74), which indicates that CoProStory's design affects participants' collaborative behaviors and children would show more positive reactions in version 2. The average negative data (C2) are 0.14(SD=0.35) and 0(SD=0). The negative data are both low, which indicates that children would unlikely show negative reactions while programming with CoProStory.

For helping behaviors (C3), children who used G1 showed relatively fewer helping behaviors ( $M=3.29$ ,  $SD=3.1$ ) compared to children who used G2 ( $M=5.29$ ,  $SD=2.12$ ). And for C4, children in G2 were more likely to ask for help ( $M=0.86$ ,  $SD=0.99$ ) than G1 ( $M=1.71$ ,  $SD=2.19$ ). The results of G2 were relatively higher than G1 which means the design of Sync Blocks can promote children’s helping behaviors. Combined video data, we can find that children are more likely to help their partners while programming with Sync Blocks. Here is an example in G2:

**Children1:** Now I need the suspension bridge down to continue. Using Wait Blocks to wait...where do you get? I need you place the Notice Blocks.

**Children2:** I haven’t got there. And I need the climb Animation Block.

**Children1:** Here is your Block. What’s next?

**Children2:** Let’s see...

As the example shows, children in G2 were more likely to pay attention to their partners with Sync Blocks while programming. By contrast, according to the video data, children in G1 often offered initiative help until they finished their own tasks. Here is an example in G1:

**Children1:** I have finished my part!

**Children2:** I still need...

**Children1:** It’s easy. Let me help you.

The results of our interview are shown in Figure 6. Overall, scores of interviews show that CoProStory could support children programming collaboratively.

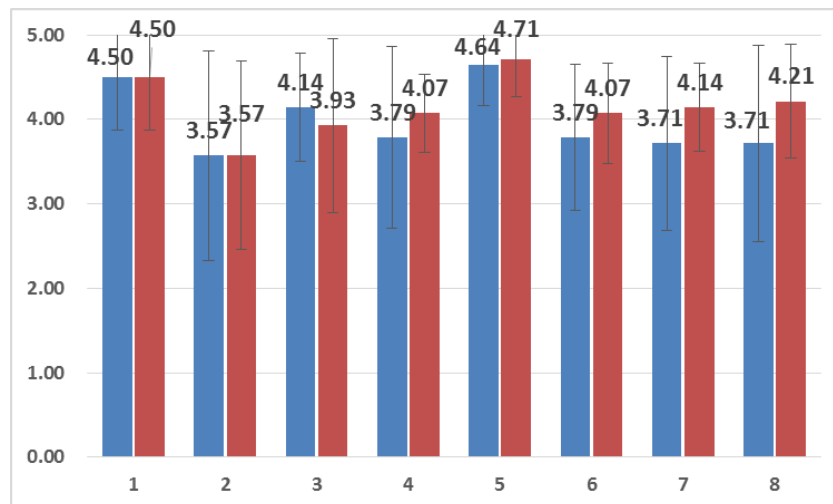


Figure 6. Scores of Q1-Q8 (blue columns are scores of G1, red columns are scores of G2).

The result of the interview shows that CoProStory is interesting, easy to use for children for the reason that children in both G1 and G2 gave CoProStory average 4.5( $SD=0.63$ ). But the scores of Q2, both G1 and G2 result in 3.57, indicate that we may make the semantics of the blocks clearer to reduce the cognitive load. As for Q3, the average scores (4.14 with  $SD=0.64$  and 3.93 with  $SD=1.03$ ) of two groups were positive which means children thought it is easy to use programming blocks to control the characters. Q4 is about whether the real-time feedback in the game helpful for children in the programming stage. Children in different groups gave average scores of 3.79( $SD=1.08$ ) and 4.07( $SD=0.46$ ). The result shows a positive evaluation of CoProStory’s real-time feedback and indicates that the real-time feedback was helpful for children to complete the tasks.

In Q5, children were asked how much they like playing the game with their partner, they gave the question average 4.64 ( $SD=0.48$ ) and 4.71( $SD=0.45$ ). Both groups gave high scores. In Q6, we asked the children if they paid attention to their partner when they were playing the game. They gave the question average 3.79 (G1,  $SD=0.86$ ) and 4.07(G2,  $SD=0.59$ ). A relative higher score of G2 indicates that children may pay more attention to

their partner with Sync Blocks. In Q7, we investigated how much help did they offered their partner, children gave an average of 3.71 (G1, SD=1.03) and 4.1(G2, SD=0.52). Children also gave an average of 3.71(G1, SD=1.16) and 4.21(G2, SD=0.67) for Q8-investigating how much help their partner offered them and the difference between G1 and G2 indicates that Sync Blocks may lead children help each other.

## Discussion

CoProStory is a new tangible programming tool for children's collaboration and the design of Sync Blocks is to promote collaboration and avoid inequality of loads of programming. According to the results of user study, children were more engaging and showed more positive reactions while programming with Sync Blocks. But as the results of interview shows, there isn't significant difference in the scores of interests between these 2 versions and children gave lower scores to learn in G2. The reason may be that the new Sync Blocks also increase the cognitive load of programming while coordinating children's programming.

To avoid inequality of loads, we make clear role division and assign tasks to children as expected. However, the separation may cause less collaboration and the difference of programming speed may still cause inequality, which is reflected in G1: children in G1 often offered initiative help until they finished their own tasks and as consequence the children got help might take less loads. To avoid inequality and promote collaboration, we use Sync Blocks to make the separated tasks time-related and coordinate children's programming. And based on our experiment, we think it's necessary to classify the help behaviors and find new measurement of inequality to get a better comparison. Besides, the design of visual games with Sync Blocks needs further consideration. For instance, the positions of time-related nodes effect children's collaboration and the method of nodes assignment needs further research.

## Conclusion

This paper described CoProStory, a brand-new tangible programming tool which could encourage collaboration between children. The tool supports children to program together in the same physical environment while work on a shared quest. Besides, with the comparison of the two versions, we found the advantage of the design of Sync Blocks.

We try to applied the concept of process synchronization to promote children's collaboration. As described, we design Sync Blocks to coordinate children's programming and made a contrast user study to explore the better way to support children's collaboration better. And as the results of our experiment show, children were more positive in version 2. In particular, some children who program faster usually would help their partners if their program was blocked by Wait Blocks due to the turning of attention to their partners. Besides, children would observe other programming while using Sync Blocks, which usually indicates positive discussions.

In the future, this work could be improved in several ways. As the results of the interview show, we need to improve the interaction of CoProStory. For instance, we can make our real-time feedback clearer. Furthermore, the interplay between quality of collaboration and learning outcome is worth to be further studied in depth through designs of more specific learning modules and corresponding evaluation methods.

## References

- Andy Cockburn, Andrew Bryant. (1997). Leogo: An equal opportunity user interface for programming: *Journal of Visual Languages & Computing* 8, 5: 601-619.
- Charlie McDowell, Linda Werner, Heather Bullock and Julian Fernald. (2002). The Effects of Pair-Programming on Performance in an Introductory Programming Course. In *Proceedings of the SIGCSE technical symposium on Computer science education (SIGCSE'02)*, 38-42.
- Douglas H. Clements. (1999). The future of educational computing research: the case of computer programming. *Information Technology in Childhood Education Annual*, 147-179.
- Florian Scharf, Thomas Winkler, Claudia Hahn, Christian Wolters and Michael Herczeg. (2012). Tangicons 3.0: an educational non-competitive collaborative game. In *Proceedings of the 11th International Conference on Interaction Design and Children*, 144-151.
- Hideyuki Suzuki and Hiroshi Kato. (1995). Interaction-Level Support for Collaborative Learning: AlgoBlock—an open programming language. In *Proceedings of the first international conference on Computer support for collaborative learning*, 349-355.
- Hiroshi Ishii, Brygg Ullmer. (1997). Tangible bits: towards seamless interfaces between people, bits and atoms. In *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*, 234-241.

- Hyunjoo Oh, Anisha Deshmane, Feiran Li, Ji Yeon Han, Matt Stewart, Michael Tsai, Xing Xu, and Ian Oakley. (2013). The Digital Dream Lab: Tabletop Puzzle Blocks for Exploring Programmatic Concepts. In Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction, 51-56.
- Jeannette M. Wing. (2006). Computational thinking. *Communications of the ACM* 49, 3: 33-35.
- Leslie Eugene Chipman. (2007). Collaborative technology for young children's outdoor education[M]. ProQuest.
- Michael S. Horn, R. Jordan Crouser, Marina U. Bers. (2012). Tangible interaction and learning: the case for a hybrid approach. *Personal and Ubiquitous Computing* 16, 4: 379-389.
- Michael S. Horn, Robert J. K. Jacob. (2007). Tangible Programming in the Classroom with Tern. In Proceedings of the CHI '07 Extended Abstracts on Human Factors in Computing Systems.
- Neha Katira, Laurie Williams, Eric Wiebe, Carol Miller, Suzanne Balik and Ed Gehringer. (2004) . On understanding Compatibility of Student Pair Programmers. In Proceedings of the SIGCSE technical symposium on Computer science education (SIGCSE'04), 7–11.
- Regan L. Mandryk, Kori M. Inkpen, Mark Bilezikjian, Scott R. Klemmer and James A. Landay. (2001). Supporting children's collaboration across handheld computers. In Proceedings of the ACM SIGCHI Conference on Human factors in computing systems, 255-256.
- Robert F. Bales. (1950). A set of categories for the analysis of small group interaction. *American Sociological Review*, 257–263.
- Scott S D, Regan L. Mandryk, Kori M. Inkpen. (2003). Understanding children's collaborative interactions in shared environments[J]. *Journal of Computer Assisted Learning*, 19(2): 220-228.
- Timothy S. McNerney. (2004). From turtles to Tangible Programming Bricks: explorations in physical language design. *Personal and Ubiquitous Computing*, Volume 8, 326-337.
- Tomohito Yashiro, Yasushi Harada, and Kazushi Mukaiyama. (2017). Plugramming: A Tangible Programming Tool for Children's Collaborative Learning. *International Conference on Human-Computer Interaction*, 398-409.

## Acknowledgments

This research is supported by the National Key Research and Development Program under Grant No. 2016YFB0401202, and the National Natural Science Foundation of China under Grant No. 61872363, 61672507, 61272325, 61501463 and 61562063.

# A Spiral Model of Collaborative Knowledge Improvement to Support Collaborative Argumentation for Science Learning: Technological and Pedagogical Design

Wenli Chen, Si Zhang, Yun Wen, Chee-Kit Looi, and Jennifer Yeo  
wenli.chen@nie.edu.sg, si.zhang@nie.edu.sg, yun.wen@nie.edu.sg, cheekit.looi@nie.edu.sg,  
jennifer.yeo@nie.edu.sg  
Nanyang Technological University, Singapore

**Abstract:** Innovations in teaching and learning are not merely about the design of technologies but the integration of technologies and pedagogical practices in supporting meaningful learning. This paper presents the design, implementation and evaluation of a web-based system to support secondary school students' collaborative argumentation (CA) in science learning in Singapore. A pedagogical model named the Spiral Model of Collaborative Knowledge Improvement (SMCKI) is proposed in this study to inform the system and learning activity design. Starting with a stage of individual brainstorming, the pedagogical model scaffolds students to go through five stages of intra-group and inter-group knowledge improvement and refinement, to support the advancement of their collective and individual knowledge. The results showed that the students significantly improved on their scientific content knowledge through the staged collaboration argumentation activities in the web based learning environment.

## Introduction

In recent years there is a shift in science education from focusing on exploration and experiment to the construction of argument and explanation (Duschl & Osborne, 2002). Argumentation refers to the process of discussion and negotiation among peoples of different point of view (Osborne, Erduran, & Simon, 2004; Sampson & Clark, 2009). Argumentation is part of the practice of science for evaluating, refining and establishing new theories (Duschl & Osborne, 2002). It has been widely recognized as an effective approach for science learning (Osborne & Patterson, 2011) as it helps students improve their conceptual understanding (Bouyias & Demetriadis, 2012), understand the nature of science, promotes deeper learning of content (Nussbaum, 2008), enhance knowledge creation (Erduran, Simon, & Osborne, 2004), and develop metacognitive skill (Böttcher & Meisert, 2011).

Many effective argumentation happens among students (Scheuer, Loll, Pinkwart, & McLaren, 2010) who engage in proposing, critiquing, coordinating evidence with claims to construct arguments and explanations, reflecting, and evaluating each other's' ideas. Educational researchers have developed some pedagogical approaches and tools to support students' collaborative argumentation (Scheuer et al., 2010). However, students were often found as working ineffectively and inefficiently when doing the argumentation. Students are still not substantively engaged in the process of discussion and negotiation (Yun & Kim, 2015). One of the critical issues is that students' discussions do not lead to significant improvement of idea improvement due to the lack of interdependence among group members. More carefully designed collaborative argumentation activities which can support idea improvement is needed.

This paper presents a web-based platform to support students' collaborative argumentation in science learning. Developed by National Institute of Education, Nanyang Technological University Singapore, this system is designed for supporting generalized coordination of collaborative argumentation among students and the teacher in the following three aspects: 1) argumentation: developing graph-based argumentation to represent argument elements, and relationships between them, 2) collaboration: scaffolding student's continuous knowledge improvement through a staged Spiral Model for Collaborative Knowledge improvement, 3) peer assessment and critique: supporting intra-group peer assessment and critique by quantitative rating and qualitative feedbacks.

## Literature review

### Argumentation and learning

Argumentation is viewed as a vital type of knowledge construction activity that can lead to knowledge advancement (Weinberger & Fischer, 2006). Science teachers have designed various learning activities to support students' argumentation (Emig, McDonald, Zembal-Saul, & Strauss, 2014), and students are encouraged to participate in scientific practice activities, reflective dialogue, and actively use evidence to support their own



claims (Bulgren et al., 2014; Duschl et al., 2002). Argumentation is as an effective learning process and educational outcome for science education (Osborne & Patterson, 2011).

Many argumentative frameworks have been proposed to support students' argumentation in science education. Toulmin Argument Pattern (TAP) is the most commonly used framework (Toulmin, 1958). TAP consists of six elements: claim, qualifier, data, warrant, backing, and rebuttal. A *claim* is an assertion, or statement, about a belief or idea. *Grounds* are statements or reasons that support the *claim*. *Warrants* are an elaboration on the reasoning behind why the person believes their *claim* to be true. A *qualifier* provides strength and clarification to the *grounds* and *warrant*. With a *qualifier*, the claim is valid only during a specific circumstance. A *rebuttal* is a particular condition in which the *warrant* becomes void and the *claim* is not valid. While a *qualifier* can provide strength and clarification, the *backing* provides support to the *warrant* by stating why the *warrant* is acceptable.

Another argumentation framework is Walton's argumentation scheme (Walton & Reed, 2005). Walton breaks down argumentation into six different types of dialogue: persuasion dialogue, the inquiry, negotiation, information-seeking dialogue, deliberation, and eristic dialogue. Persuasion dialogue is when one person is trying to persuade another person that some particular suggestion is true through the use of arguments that show or prove that it is true. Inquiry dialogue involves a group of people investigating the reasons for some event or phenomenon. Negotiation dialogue involves bargaining, while information-seeking dialogue is when one person has, or appears to have, information that another person wants. Deliberation dialogue is making a decision to solve a problem and lastly, eristic dialogue is a quarrel between two people. The argumentative patterns are widely used to evaluate students' argumentation ability and its development (Garcia-Mila, Gilabert, Erduran, & Felton, 2013).

In addition to providing argumentative frameworks to help students develop their ability to argue, many studies have found that graph-based computer application tools can help students develop argumentation skills and scientific knowledge (Dwyer, Hogan, & Stewart, 2012; Hsu, Dyke, Smith, & Looi, 2018). Compared with these previous studies, this study especially emphasizes the integration of technology and pedagogy. While technical tools support scientific argumentation, pedagogical design better supports students' collaborative learning.

## Collaborative argumentation

As reflected in literature, students encounter a lot of difficulties in collaborative argumentation. Except for unequal participation and low level engagement as often observed in collaborative learning activities (Choo, Eshaq, Samsudin, & Guru, 2009), students might also not use appropriate and/or enough evidence to warrant their claims, or justify their choice of evidence in the arguments they produce (Erduran, Simon, & Osborne, 2004; Weinberger, Stegmann, & Fischer, 2010). The self-explanations made during argumentation were only sporadically generated by even good students (Larsson & Alterman, 2009). Students did not evaluate the validity or acceptability of an explanation for a given phenomenon in a satisfying way as well (Bouyias & Demetriadis, 2012). There was also evidence that students had problem producing "a good argumentation sequence" (Larsson, & Alterman, 2009). When required to engage in scientific argumentation, a critical type of discourse epistemic practice in doing and communicating science (Lazarou, Sutherland, & Erduran, 2016), students were confronted with more challenges. Existing research has indicated that students often lacked ability to determine the acceptance, rejection, or modification of ideas (Hogan & Maglienti, 2001). Their acts of distorting, trivializing, and ignoring certain evidence frequently resulted in misconceptions (Sampson & Clark, 2008). In the realm of CSCL research, collaborative argumentation is regarded as a key type of knowledge construction process that should be mastered by students to enable knowledge advancement. Yet in the argumentation processes, students without necessary training were often noted as working ineffectively and inefficiently. They did not ask each other questions, not explain or clarify their own opinions, not articulate the reasoning behind, not elaborate and reflect on knowledge (Kollar, Fischer, & Hesse, 2006; Martin & Hand, 2009). They also hardly take alternative perspectives into consideration (Sampson, Grooms, & Walker, 2011).

In science education, engaging in collaborative argumentation is challenging for most students (Gould & Parekh, 2018). During the process of collaborative argumentation, it is difficult for students to construct scientific explanations, and students adopt the wrong ways of arguing and only a few students can complete the argumentation task (Ryu & Sandoval, 2012). The main reason for this phenomenon is that students lack the basic skills of collaborative argumentation. Students don't know how to create a logical point of view and present evidences for support it. To improve students' performance of collaborative argumentation, one kind of method is to provide students with a structured argumentation scaffold (Suthers, 2003).

## Web-based system design

The web-based system development is based on the J2EE platform, and adopts the SSH (Spring, Struts, and Hibernate) technology. The system includes three modules: graph-based argumentation, collaborative knowledge



improvement, and peer assessment and critique. Figure 1 shows the screenshots of the system. The central area of the screen is students' graph-based argument workspace, where students use evidence to support or oppose the claim. The activity description (such as activity topic, activity introduction, role assignment, experimental data and evaluation rules, etc) and chat window are shown on the top right of the page.

## Graph-based argumentation design

The system is a graph-based argumentation application that uses different shapes to represent the argument elements. The oval node represents the claim, the cloud node represents the idea (not sure if it is claim or evidence), and the rectangle node represents the evidence. The explanation of shapes of the argument elements in the system are shown in figure 1. The arrows represent the relationships among the nodes. Green arrows represent "evidence for" whereas red arrows represent "evidence against".

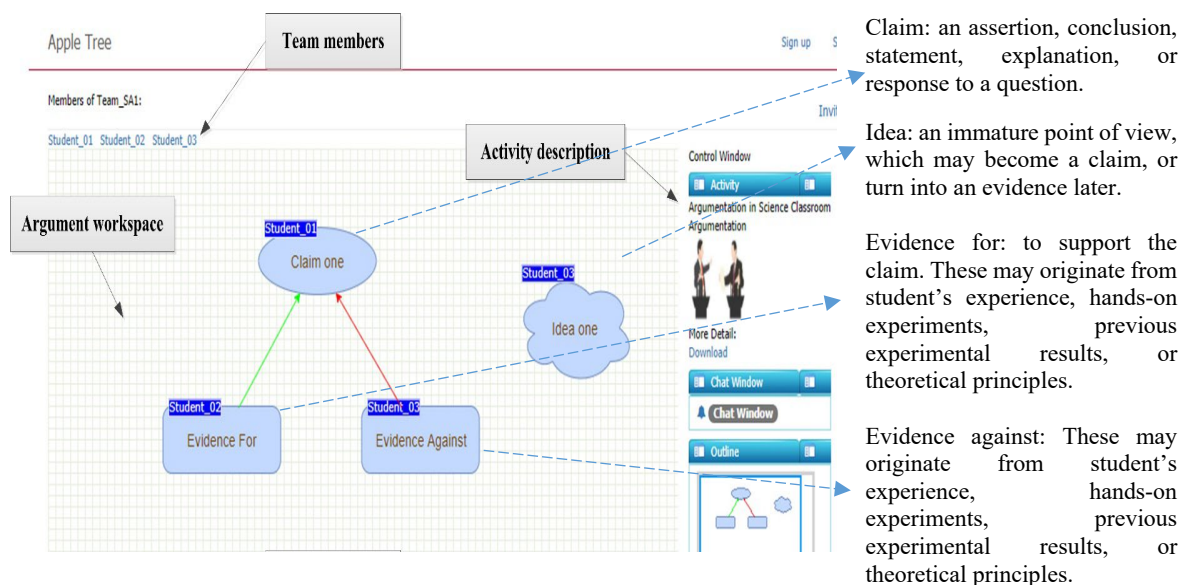


Figure 1. The Screenshot and explanation of argument elements of the System.

The system supports uploading and downloading attachments (including pictures and documents) with the nodes. In addition, the system supports the transitions between different shapes, for example, the idea can be changed to the claim, or changed to the evidence if it was connected with the claim. Because the team members are engaged in collaborative argumentation, members from the same group can modify or delete the nodes/relationships created by other members of the group.

## Collaboration design: Spiral Model of Collaborative Knowledge Improvement (SMCKI)

Instead of collaboration scripts that generally provide a detailed set of guidelines, rules and structured tools for describing how the group members should interact, we propose a pedagogical model to guide the interaction processes so that Collaborative Knowledge Improvement can happen more effectively in classrooms. The 5-stage Spiral Model of Collaborative Knowledge Improvement (SMCKI) provides a tangible structure for one operational collaborative activity design beginning with brainstorming and a structured process of constant knowledge improvement. The model focuses on democratic knowledge sharing as well as cycles of individual, group and class knowledge enhancement (see Figure 2).

This model is based on the authors' previous work on Funnel Model for rapid collaborative knowledge improvement (Wen, Looi, & Chen, 2011), which consists of three stages of collaborative learning process. By respecting and encouraging cognitive diversity, the first stage encourages the creation of diverse ideas. The subsequent stages tap on this diversity to seek synergy of ideas, and a stage of convergence and consensus seeking leading to knowledge convergence and advancement of the individuals, groups and class. The SMCKI model entails 5 stages (Figure 2). The teacher used verbal instructions and system buttons to regulate the 5 stages of scientific argumentation activity.

I: Individual brainstorming: Students individually construct argument with claims and evidences of the scientific phenomena. The argument represents the best knowledge of the individuals.

II: Intra-group synergizing: Students discuss, synergize and consolidate group members' work by deleting, adding, modifying nodes/relationships. A group graph-based argumentation diagram is created which represents the best knowledge of the group.

III: Inter-group peer assessment and critique: Students go to other groups to provide quantitative ratings and qualitative comments by identifying the strengths and areas for improvements.

IV: Intra-group refinement: Students go back to their own groups and refine the group work based on other groups' ratings and feedbacks. After further verbal negotiation, they were required to seek consensus and finalize their group idea.

V: Individual idea perfection: Individually, students write an argumentative essay or reflection report to explain the scientific phenomena.

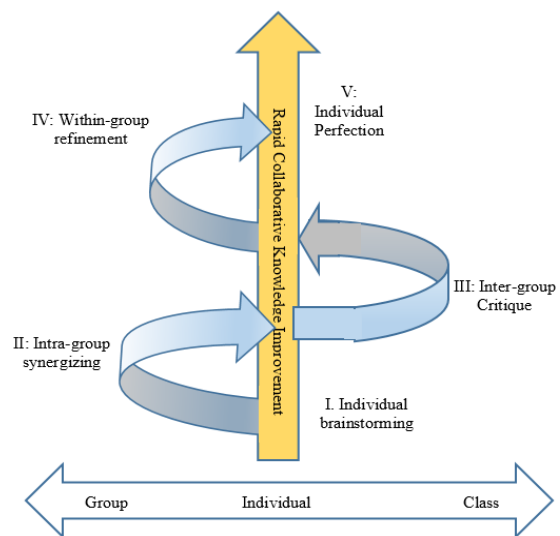


Figure 2. Spiral Model of Collaborative Knowledge Improvement.

### Peer assessment and critique

In stage III of SMCKI, students go to other groups to provide quantitative rating and qualitative feedbacks on the content and structure of the argument. Students can rate both the nodes and the relationships in the argument diagram. When students double-click the node or relationship, an assessment window will pop up (see Figure 3). All the students' qualitative feedbacks will be displayed in a table format. In order to encourage thoughtful critique and rating, each element of the argument diagram can only be rated once. Students name will not be shown in this stage.

To make the ratings more intuitive for students, the color of the lines of those nodes and relations which received positive ratings will be highlighted whereas the color of the lines of those nodes and relations which received negative ratings will be lighter. This feature will help channel teacher and students' attention when they do stage 3 and 4 argumentation activities. On top of assessing the individual elements of the argument diagram, students can assess the whole diagram based on the rubrics given by the teacher.

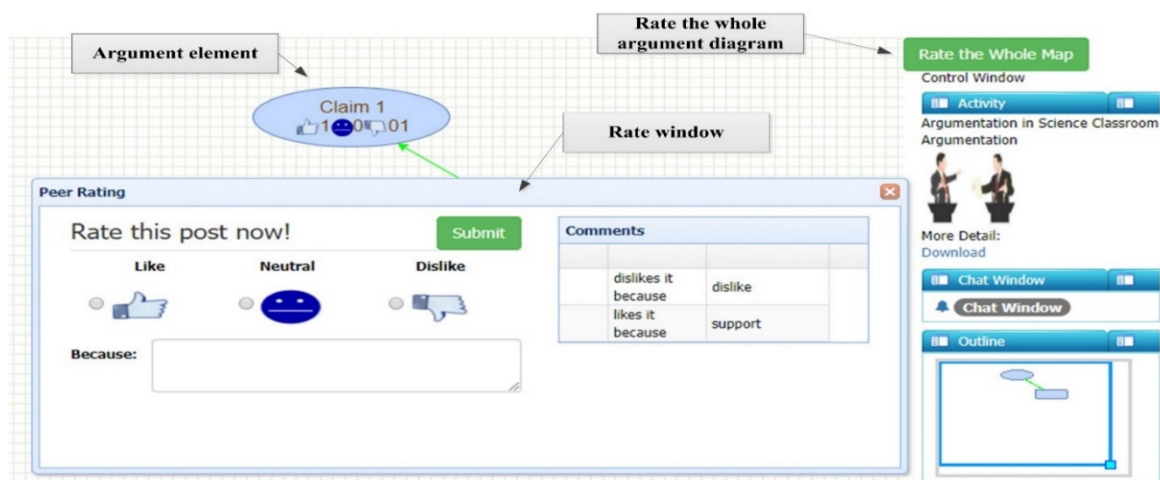


Figure 3. The screen shot of the peer assessment and critique (Stage III of SMCKI).

### Research questions

In this study, we focus on investigating whether and how was the use of web-based system informed by the SMCKI helpful for improving students' conceptual knowledge learning?

## The enactment of SMCKI in science classroom

### Participants

A secondary grade 4 (16 years old on average) class with a total of 33 students in Singapore participated in the project. They studied physics phenomenon through SMCKI approach using the system. All the students in this school are equipped with a personal iPad 24/7, and they are proficient in using iPad as a learning tool. The teacher has rich experiences in science teaching and he is tech-savvy. Before the project started, researchers conducted training for the teacher with the aim to improve teacher's understandings of the technological and pedagogical design of the system. All the students were heterogeneously grouped by the teacher according to their ability. There were nine groups of 3-4 students in the class.

The collaborative argumentation activities were co-designed by the teachers and researchers. Three lessons were implemented (50 minutes per lesson) from July to August in 2018. The data analyzed in this paper were from the third lesson on the electromagnetic induction phenomenon. Following the SMCKI approach, students went through 5 staged of collaborative argumentation. The time allocation for each stage is: Individual brainstorming (7 minutes), within-group synergizing (10 minutes), inter-group assessment and critique (8 minutes), intra-group refinement (10 minutes), and individual perfection (after class). The paper examines the students' conceptual development in the process for 5-staged collaborative argumentation activities.

### Methods and instruments

During the collaborative argumentation activities, students were asked to provide explanation about the phenomena that they observed from an iPad which showed the induced current flowed in a solenoid over time when a magnet fell through it. Students presented their explanations based on an explanation framework which consisted of three elements: claim, evidence, and reasoning (McNeill, Lizotte, Krajcik, & Marx, 2006). A total of 48 explanations generated in students' collaborative argumentation activities. Content analysis was employed to analyze the nature of peer comments in Stage 3. The unit of content analysis is one comment.

*Pre-test and post-test design.* Pretest and posttest were used to measure students' conceptual knowledge before and after the collaborative learning activity and used the same text paper, which contained 4 questions, one point for each question. The test questions were closely related to the knowledge points in this lesson.

*Coding scheme of peer feedbacks.* A coding scheme adapted from Clark & Sampson (2007) was used to code students' qualitative feedbacks in stage 3. The coding scheme consists of five codes: rebuttal, support, query, emotive appeal, and off-task comments. The second and third authors independently coded all the 79 comments, with an inter-rater reliability of 0.864 (Cohen's Kappa).

*Students' interview design.* Students were required to reflect on their experiences of collaborative learning based on one guiding questions in the post-intervention interview: What did you do at each stage of collaborative learning activity?

### Data analyses and results

#### Comparison of students' conceptual understanding between Stage I and Stage V.

In Stage I and V, individual students are required to construct argument and explanation of the scientific phenomenon individually. A comparison was made to measure students' conceptual knowledge between stage I and V. Descriptive statistics were used to analyze the mean and standard deviation of conceptual understanding, see Table 1. As the data do not conform to normal distribution, the Wilcoxon signed Ranks test was used to detect differences on conceptual knowledge between 2 stages. The results in table 1 showed that students' conceptual understanding in Stage V was significantly higher than the stage I ( $z=-4.647^{**}$ ,  $p<0.001$ ).

Table 1: Wilcoxon signed Ranks test results

|         | N  | Minimum | Maximum | Mean | Std.Dev | Z         |
|---------|----|---------|---------|------|---------|-----------|
| Stage I | 33 | 0       | 4       | 2.58 | 0.830   | -4.647*** |
| Stage V | 33 | 2       | 4       | 3.70 | 0.529   |           |

Note. \*\*\*  $p<0.001$ .

#### Students' conceptual development throughout staged activities: Group 3 as the case study

We provide a qualitative account of group 3's process of doing the staged collaborative argumentation. In order to clearly show the process of students' revision of their explanations, we used bold font and underlining in the text to show the position of students' revision.

*Stage I individual posting.* All 4 students (Jenny, Hellen, Nichole, and Sabrina, all pseudo names) managed to post their explanation of the scientific phenomena based on their pre-existing knowledge.

*Stage II intra-group synergizing.* After seeing each other's posting, students compared each other's explanations and tried to synergize their work through negotiation and discussion. In this stage, Helen added a scientific knowledge element to her previous explanation in stage 1.

*“When the magnet first enters the solenoid, there is a momentary increase in rate of cutting of magnetic field lines by the coil, hence by Faraday’s Law, an e.m.f. is induced in the coil and there is a current. In the moment that the magnet is completely in the solenoid, there is no rate of cutting of magnetic field lines by the coil, so there is no e.m.f. induced in that moment and the current is 0. When the magnet falls out of the solenoid, there is a momentary decrease in rate of cutting of magnetic field lines by the coil, hence by Faraday’s Law, an e.m.f. is induced in the coil. Since the motion of the magnet is in the opposite direction as compared to when it enters the solenoid, the induced e.m.f. is in the opposite direction and the current is also in the opposite direction, **by Lenz’s Law.**” (14:02, posted by Helen)*

In this stage, Sabrina repeatedly revised the statements that represent her explanation. In the process of revision, the representation of Sabrina's explanation moved closer to the expert scientific concept knowledge.

*“When the magnet is dropped into the solenoid, **there is a momentary rate of cutting of magnetic field lines and thus an e.m.f is induced and a current in the solenoid is recorded.** According to Lenz's law of motion, as the magnet changes direction to move out of the magnet, the current recorded would be of **similar magnitude** and amplitude but in a different direction.” (14:06, posted by Sabrina)*

In stage II, students not only revised their own explanations, but also added new explanations. Sabrina added a new claim and evidence to improve own explanation. Jenny put forward a backing to Helen's explanation. During the post-activity interviews, the 4 students explained to researcher how they synergized their group argumentation diagram in Stage 2. Students discussed with her group mates first and then modified her explanation (from the interview transcript with Sabrina). If student's concept is wrong, she would modify her explanation (from the interview transcript with Sabrina). If students thought their partners' explanations make more sense, they will add others' idea to her explanation (from the interview transcript with Helen).

*Stage III Intra-group assessment and critique.* Group 4 students went to group 3 to provide ratings and feedbacks. To examine the role of stage 3 peer critique for students knowledge improvement in argumentation, content analyses was done to understand the nature of peer comment given. Group 4 gave a total of 7 comments in this stage, including 1 support, 3 rebuttals, and 3 emotive appeals. Table 2 shows an example of comments given by group 4.

Table 2: An example of comments given by group 4 to group 3

|                       |  |
|-----------------------|--|
| 1.Explanation         | Group 3 (Student <i>Jenny</i> )<br>“When the magnet is drops into the coil, there is a momentary rate of cutting of the magnetic field lines from the magnet by the coil. By faraday’s law, an emf is induced in the coil, which drives an induced current, hence a sudden increase in magnitude of current was recorded. This repeated in the opposite direction when the magnet fell into the solenoid, hence by lenz’s Law, the direction of the current is opposite to that at first, so a current of similar magnitude but in opposite direction was recorded.” |
| 1.1<br>Rebuttal       | Group 4 (comments to group 3)<br>“why does the direction change? ”   |
| 1.2<br>Rebuttal       | Group 4 (comments to group 3)<br>“What does “direction of current is opposite to that at first” a bit confusing.”  |
| 1.3<br>Emotive appeal | Group 4 (comments to group 3)<br>“True”  |

During the post-activity interviews, students explained their experiences in this stage. In this stage, students of group 4 gave comments to group 3 based on the following criteria: (1) key words, (2) explanation was clear or not, (3) the number of evidence supported one claim, (4) students' understanding of the conceptual knowledge, and (5) the concept was correct or accurate. If students thought that the other group's explanations were better, they would bring other groups' ideas to their own group.

*Stage IV Intragroup refinement.* Based on the explanations and comments from other group, students in group 3 further revised their explanations. Students added scientific conceptual knowledge in their explanations or had their doubts about some conceptual knowledge cleared.

## Discussion and conclusion

This study proposes web-based platform to support students' collaborative argumentation in science education by scaffolding students' graph based argumentation informed by a spiral model of collaborative knowledge improvement. An empirical study was conducted to investigate how student construct argumentation for science learning throughout the stage-collaborative argumentation activity. The paper provides a descriptive account of the interactions of a group as the students engaged in spiral cycles of knowledge improvement. The case study demonstrates that students' conceptual understanding improved continuously through the 5 stages of SMCKI. This study also showed that revision, deletion and synthesis of explanation were sophisticated skills, and the web-based system needed to provide scaffoldings for the development of students' these skills. In addition, teachers' feedback was also very important. In the second and fourth stages, students would perform better if teachers could provide some feedback on scientific knowledge and collaboration process.

Informed by SMCKI, our web-based system has 3 characteristics: "low floor", "wide walls", and "high ceiling". "Low floor" means all students can easily participate and contribute to the group and class's knowledge improvement; "Wide walls" means the networked online system enables everyone to connected through individual-group-class-group-individual interaction pattern. There are many opportunities for students be exposed to diversified ideas and perspectives. "High ceiling" means that students can tap on each other's ideas to improve the knowledge throughout the staged spiral model. Our empirical study shows that the quality of ideas of the one stage is always higher than the preceding stage. Students work on increasingly complicated ideas by communication, negotiating and critique, all of which require higher order thinking.

The SMCKI is also meant to scaffold teachers to enact and orchestrate the collaborative learning activities in the classroom as well as to build capacity to be able to design such learning activities themselves. It has the potential for teachers to embark on collaborative learning activities in the classroom and to manage the risks of the activities breaking down or not reaching any kind of fruitful collaboration. SMCKI is not specific to collaborative argumentation or science learning. It can be applied in any collaborative learning activity. Further study is needed to investigate the application of SMCKI in other contexts of collaborative learning.

## References

- Bouyias, Y., & Demetriadis, S. (2012). Peer-monitoring vs. micro-script fading for enhancing knowledge acquisition when learning in computer-supported argumentation environments. *Computers & Education*, 59(2), 236-249.
- Böttcher, F., & Meisert, A. (2011). Argumentation in science education: A model-based framework. *Science & Education*, 20, 103-140.
- Bulgren, J. A., Ellis, J. D., & Marquis, J. G. (2014). The use and effectiveness of an argumentation and evaluation intervention in science classes. *Journal of Science Education and Technology*, 23(1), 82-97.
- Choo, K. A., Eshaq, A. R. M., Samsudin, K. A., & Guru, B. K. (2009). An evaluation of a constructivist online collaborative learning activity: A case study on Geometry. *Turkish Online Journal of Educational Technology*, 8(1), 15-25.
- Clark, D. B., Sampson, V., Weinberger, A., & Erkens, G. (2007). Analytic frameworks for assessing dialogic argumentation in online learning environments. *Educational Psychology Review*, 19(3), 343-374.
- Duschl, R. A., & Osborne, J. (2002). Supporting and promoting argumentation discourse in science education. *Studies in Science Education*, 38, 39-72.
- Dwyer, C. P., Hogan, M. J., & Stewart, I. (2012). An evaluation of argument mapping as a method of enhancing critical thinking performance in e-learning environments. *Metacognition and Learning*, 7(3), 219-244.
- Emig, B. R., McDONALD, S. C. O. T. T., ZEMBAL - SAUL, C. A. R. L. A., & Strauss, S. G. (2014). Inviting argument by analogy: Analogical - mapping - based comparison activities as a scaffold for small - group argumentation. *Science Education*, 98(2), 243-268.
- Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: Developments in the application of Toulmin's argument pattern for studying science discourse. *Science education*, 88(6), 915-933.

- Garcia-Mila, M., Gilabert, S., Erduran, S., & Felton, M. (2013). The effect of argumentative task goal on the quality of argumentative discourse. *Science Education*, 97(4), 497-523.
- Gould, D. L., & Parekh, P. (2018). Mentoring and Argumentation in a Game-Infused Science Curriculum. *Journal of Science Education and Technology*, 27(2), 188-203.
- Hogan, K., & Maglienti, M. (2001). Comparing the epistemological underpinnings of students' and scientists' reasoning about conclusions. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, 38(6), 663-687.
- Hsu, P.-S., Dyke, M. V., Smith, T. J., & Looi, C.-K. (2018). Argue like a scientist with technology: the effect of within-gender versus cross-gender team argumentation on science knowledge and argumentation skills among middle-level students. *Educational technology research and development*, 1-34.
- Kollar, I., Fischer, F., & Hesse, F. W. (2006). Collaboration scripts—a conceptual analysis. *Educational Psychology Review*, 18(2), 159-185.
- Larusson, J. A., & Alterman, R. (2009). Wikis to support the “collaborative” part of collaborative learning. *International Journal of Computer-Supported Collaborative Learning*, 4(4), 371-402.
- Lazarou, D., Erduran, S., & Sutherland, R. (2017). Argumentation in science education as an evolving concept: Following the object of activity. *Learning, Culture and Social Interaction*, 14, 51-66.
- Martin, A. M., & Hand, B. (2009). Factors affecting the implementation of argument in the elementary science classroom. A longitudinal case study. *Research in Science Education*, 39(1), 17-38.
- McNeill, K. L., Lizotte, D. J., Krajcik, J., & Marx, R. W. (2006). Supporting students’ construction of scientific explanations by fading scaffolds in instructional materials. *Jrl of the Learning Sciences*, 15(2), 153-191.
- Nussbaum, E. M. (2008). Collaborative discourse, argumentation, and learning: Preface and literature review. *Contemporary Educational Psychology*, 33(3), 345-359.
- Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of research in science teaching*, 41(10), 994-1020.
- Osborne, J. F., & Patterson, A. (2011). Scientific argument and explanation: A necessary distinction? *Science Education*, 95(4), 627-638.
- Ryu, S., & Sandoval, W. A. (2012). Improvements to elementary children's epistemic understanding from sustained argumentation. *Science Education*, 96(3), 488-526.
- Sampson, V., & Clark, D. (2009). The impact of collaboration on the outcomes of scientific argumentation. *Science education*, 93(3), 448-484.
- Sampson, V., & Clark, D. B. (2008). Assessment of the ways students generate arguments in science education: Current perspectives and recommendations for future directions. *Science Education*, 92(3), 447-472.
- Sampson, V., Grooms, J., & Walker, J. P. (2011). Argument - Driven Inquiry as a way to help students learn how to participate in scientific argumentation and craft written arguments: An exploratory study. *Science Education*, 95(2), 217-257.
- Scheuer, O., Loll, F., Pinkwart, N., & McLaren, B. M. (2010). Computer-supported argumentation: A review of the state of the art. *International Journal of Computer-supported collaborative learning*, 5(1), 43-102.
- Suthers, D. D. (2003). Representational guidance for collaborative inquiry. In *Arguing to learn* (pp. 27-46). Springer, Dordrecht.
- Toulmin, S. (1958). *The uses of argument*. New York: Cambridge University Press.
- Walton, D., & Reed, C. A. (2005). *Argumentation schemes and enthymemes*. *Synthese*, 145, 339–370.
- Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & education*, 46(1), 71-95.
- Weinberger, A., Stegmann, K., & Fischer, F. (2010). Learning to argue online: Scripted groups surpass individuals (unscripted groups do not). *Computers in Human behavior*, 26(4), 506-515.
- Wen, Y., Looi, C. -K., & Chen, W. L. (2011). Towards a model for rapid collaborative knowledge improvement in classroom language learning. In H. Spada, G. Stahl, N. Miyake & N. Law (Eds.), *Proceedings of Computer Supported Collaborative Learning (CSCL) 2011 (Part 2, pp. 836-840)*. Hong Kong, China: International Society of the Learning Sciences.
- Yun, S. M., & Kim, H. B. (2015). Changes in students’ participation and small group norms in scientific argumentation. *Research in Science Education*, 45(3), 465-484.

## Acknowledgments

This study was funded by Singapore Ministry of Education (MOE) under the Education Research Funding Programme (OER 07/17 CWL) and administered by National Institute of Education (NIE), Nanyang Technological University, Singapore. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Singapore MOE and NIE.

# Flexible CSCL Orchestration Technology: Mechanisms for Elasticity and Dynamism in Pyramid Script Flows

Kalpani Manathunga and Davinia Hernández-Leo  
kalpi.nisansala@gmail.com, davinia.hernandez-leo@upf.edu  
ICT Department, Universitat Pompeu Fabra, Barcelona

**Abstract:** Flow patterns (e.g., Pyramid or Snowball) formulate good practices to script collaborative learning scenarios, which have been experimented in small-scale settings widely. Applying flow patterns on large-scale contexts present challenges to educators in terms of orchestration load. Orchestration technology can support educators to manage collaborative activities; yet existing technology do not address flexibility challenges like accommodating growing numbers of students or tolerating dynamic conditions in learning settings. We define elasticity and dynamism as two key elements in the flexibility of a script. Elasticity is related to the capacity of an orchestration technology to incorporate varying participant counts. Dynamism is the capacity to maintain a pedagogically meaningful script progression in presence of different individual behaviors. In this paper we propose flow creation and flow control mechanisms to address elasticity and dynamism in orchestration technology for Pyramid flows. These mechanisms, implemented in the PyramidApp tool, have been evaluated across four scenarios varying from small to large settings. The results show that rules enabling pyramid creation on-demand and the use of timers are useful to achieve elasticity and dynamism in the pyramid formation and progression in an automatic manner.

## Introduction

Collaboration is a coordinated process by which individuals construct and maintain shared conceptions where knowledge is co-constructed socially (Roschelle & Teasley, 1995). In collaborative learning, situations are created in which particular forms of interactions among learners are expected to occur, leading to productive learning experiences. Computer-Supported Collaborative Learning (CSCL) contributes mechanisms and technologies supporting creation of such collaborative learning situations (Dillenbourg, Sanna, & Fischer, 2009; Roschelle & Teasley, 1995). Hence, CSCL environments need to be carefully designed and implemented incorporating interaction generation and regulation mechanisms. Moreover such CSCL contexts should scaffold productive interactions and/or to facilitate activity monitoring and intervening when required (Dillenbourg et al., 2009) since free collaboration does not necessarily result in fruitful learning (Dillenbourg & Jermann, 2010).

CSCL scripting means shaping up the way that collaborations are desired to happen with technology-mediation (Dillenbourg & Tchounikine, 2007) triggering specific types of interactions beneficial for learner cognition, while achieving educational objectives. Scripts define the activity sequence, group formation, phase changing, role allocation and rotation, resource distribution, mediating communication and coordination, constrain peer interactions in social and cognitive activities that would otherwise occur rarely or not at all (Dillenbourg & Tchounikine, 2007; Kobbe et al., 2007). In this notion, scripting is possible as *micro-scripts*, emphasizing on individual learner's actions with finer granularity and *macro-scripts*, defining interactions and regulations in coarse-grained activity flows (Dillenbourg & Tchounikine, 2007; Kobbe et al., 2007). Collaborative Learning Flow Patterns (CLFPs) are examples of macro-scripts reflecting best practices to orchestrate collaborative learning which are broadly accepted and repetitively utilized by practitioners (Hernández-Leo et al., 2006). Examples of CLFPs are Jigsaw, Pyramid or Snowball, Think-Pair-Share (TPS) and Thinking Aloud Pair Problem Solving (TAPPS). Each pattern is driven by its governing pedagogy that should not be modified during the design. For example, Pyramid pattern is considered as good practice to structure collaborative learning across multiple epistemic tasks and educational levels, fostering individual accountability and positive interdependence (Davis, 2002; Hernández-Leo et al., 2006). The pedagogy of this pattern is such that individuals study a given problem initially and propose a preliminary solution. Such solutions are discussed and compared to propose a shared solution in groups. This discussing and negotiation will repeat in growing group sizes (e.g. two groups in a level of the pyramid join as a single group in the next level) until the whole group reaches consensus to propose a common solution.

Practitioners are required to invest some effort to understand pattern definitions and types of constraints to design effective, meaningful scripts (Dillenbourg & Tchounikine, 2007; Hernández-Leo et al., 2006). Moreover, in its enactment with students, they need to orchestrate or manage in real time the script mechanics (Dillenbourg & Jermann, 2010). Flexible orchestration allows to adapt CSCL scripts in real-time

with a degree of freedom to modify various orchestration aspects like group formation, role allocation and rotation (Dillenbourg & Jermann, 2010). Script modifiability is non-trivial, due to unexpected situations like learners not being present or leaving the activity in the middle spoiling on-going collaborations. As a result, orchestrating activity flows manually can be challenging for practitioners (Dillenbourg & Tchounikine, 2007; Sharples, 2013). Therefore, technology-mediated or semi-automated orchestration services are beneficial to real-time manage on-going activity flows. Previous work in the field had provided extensive knowledge in designing scripted collaborative learning flows effectively (Hernández-Leo et al., 2006; Pérez-Sanagustín, Burgos, Hernández-Leo, & Blat, 2011; Rodríguez-Triana, 2014), yet applicable mostly either at small-scale or co-located learning settings (Manathunga & Hernández-Leo, 2017). Different learning settings scaling, from small face-to-face classrooms to massive online learning communities, present varied script design and orchestration requirements. Diverse scales for collaborative learning orchestration raises a number of challenges, including adaptation challenges, tolerating unexpected conditions like drop-outs or late-joiners, needs for (re)designing scripts on-the-fly or managing the orchestration load (Sharples, 2013).

The main motivation of this research is to seek how already existing pattern-inspired scripts (in our case, Pyramid CLFP) can be enhanced to achieve flexible meaningful orchestration in order to be applied upon various learning scenarios. On the contrary to pre-defined, rigid scripts which can not be modified on-the-fly, flexible scripts allow practitioners to design and adapt in real-time with a freedom for modifications. Such modifications could be embedded during the design of the script or at the execution time as different mechanisms that do not violate the underlying pedagogical definition of the Pyramid pattern. Following sections of the article explain about those mechanisms and models introduced towards flexible CSCL orchestration, experimental settings, an analysis of the proposed mechanisms and a concluding discussion section.

## Mechanisms for flexible orchestration

The journey towards orchestration technology that supports flexible scripted collaborative learning flow is approached from two key elements that we define as *elasticity* and *dynamism*. Elasticity is defined as the capacity of an orchestration technology to accommodate growing numbers of learners to collaborative learning activities without violating the underlying pedagogical rationale of the script. Dynamism is defined as the capacity of an orchestration technology to maintain a pedagogically meaningful script progression in presence of different individual behaviors. Hence, unexpected scenarios, e.g. unanticipated activity drop-outs (Dillenbourg & Tchounikine, 2007), would not harm on-going collaborations.

As indicated above, we study the particular case of the Pyramid collaborative learning flow pattern, as an interesting structure for macro-scripts that has a potential to fit well in scenarios with a varying number of participants. Based on an analysis of the pattern structure and the targeted objectives of elasticity and dynamism, we propose a set of mechanisms named as “Flow Creation” and “Flow Control” rules. As the name implies, flow creation mechanisms are suggested at the script initiation stage, when building the flow based on its learning design, whereas flow control rules are inferred during the script execution. To show its feasibility and to evaluate the mechanisms, we have integrated their implementation into the PyramidApp tool.

## The case of the Pyramid Flow and the PyramidApp

Macro-scripts, such as those based on Jigsaw, Pyramid, Think-Pair-Share patterns, structure collaborations that potentially lead to fruitful learning (Hernández-Leo et al., 2006; Pérez-Sanagustín, Burgos, Hernández-Leo, & Blat, 2011; Rodríguez-Triana, 2014). Pyramid CLFP is structured in a way that individuals attend a given task and suggest an initial solution. Then they are assigned to small groups to discuss on the initially proposed options and agree upon a common option from the group which will be propagated to the next level(s) where much larger groups are formulated enriching collaborations and consensus reaching. At the global level, all participants agree upon one or few selected options that are reflected with the whole class. Pyramid pattern promotes individual accountability, peer interactions and positive interdependence among peers. The pattern can be applied to any subject matter, educational level and using (or not) different technologies (Hernández-Leo et al., 2006). An implementation of the pattern is reflected in PyramidApp, (Manathunga & Hernández-Leo, 2017), a web-based tool that implements an option submission space for participants to attempt the given task individually, a rating feature helping them to reach consensus with an integrated discussion space for negotiations and collaborations. PyramidApp has authoring features for the educators to create desired Pyramid pattern-inspired collaborative activities. It also provides a monitoring functionality for the educators to monitor on-going activities and to keep track of previous activities. The authoring process of PyramidApp implementation is shown in Figure 1 indicating different parameters, incorporating those related to the implementation of flow creation and control mechanisms. When enacting a PyramidApp activity, there are several phases as option submission, rating and discussion for further clarifications and to reach a consensus.



Next sections explain diverse parameters used for flow creation and control mechanisms and how those are embedded in PyramidApp.

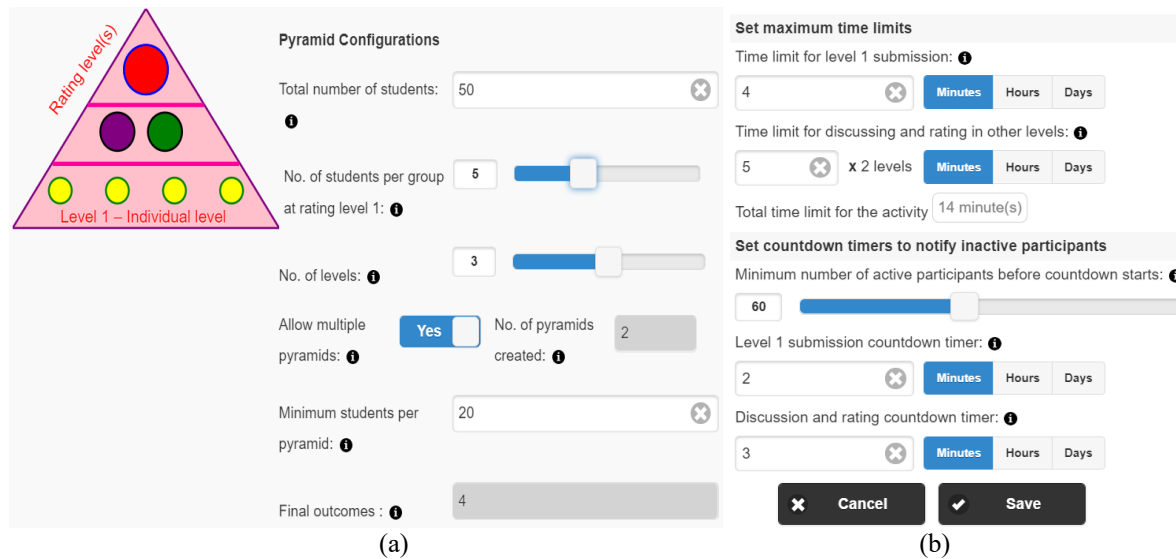


Figure 1. PyramidApp authoring application with (a) flow creation and (b) flow control mechanisms.

### Script flow creation mechanisms

As stated earlier, flow creation mechanisms were introduced to achieve elasticity, i.e. to accommodate growing numbers of learners without affecting the underlying pedagogy. As illustrated in Figure 2, the implemented rationale is creating multiple pyramids on-demand automatically and allocating students to on-going pyramids as pyramids increase sizes till a maximum threshold.

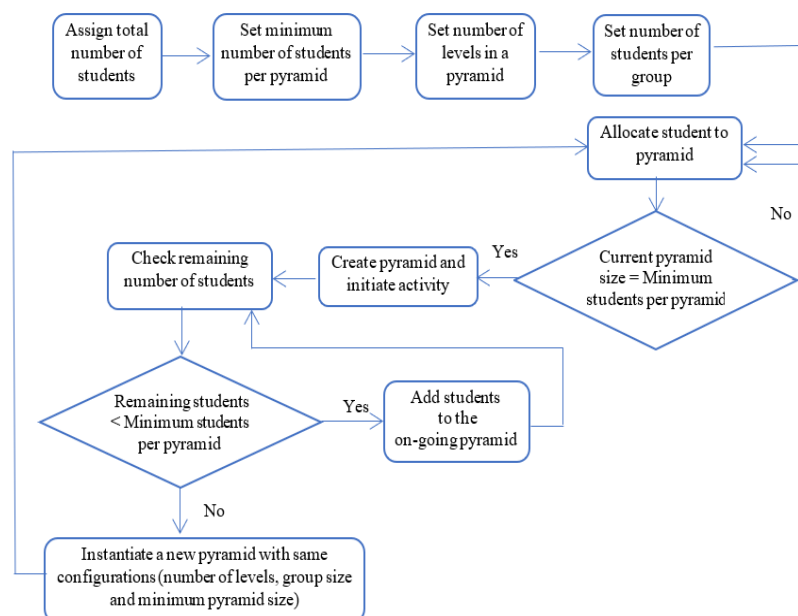


Figure 2. Flow creation mechanisms showing elastic pyramid initiation.

Key aspects introduced here are the pyramid capacity (i.e., both minimum and maximum numbers of students that can be accommodated in a single pyramid) and number of pyramids replicated to allocate further incoming students. Once a pyramid is configured by indicating number of levels in the pyramid, group size and the total number of activity participants, it grows till a maximum volume calculated using the equation:  $Max\_pyramid\_size = \{(Min\_students\_per\_pyramid * 2) - 1\}$ . When the activity is initiated, pyramids are formulated and started using the value for minimum students per pyramid and then filled till the maximum size. During the authoring phase, if a practitioner allows multiple pyramid creation (see Figure 2), PyramidApp

replicates the given design automatically to generate several pyramids to occupy the total amount of participants. When the minimum number of students per pyramid is given, the authoring tool will automatically calculate and display the possible group sizes (e.g., 2,3,4,5, etc..) along with the possible number of levels per pyramid. Based on these values and the total class size, the tool calculates the number of ultimate results from the activity after leveling through the pyramids.

### Script flow control mechanisms

Once a Pyramid flow is activated, the flow needs to be controlled with more parameters, embedding dynamic behaviors to ensure a flexible progression during the execution time. In order to achieve dynamism with a smooth pyramid progression, we introduced several timers (see Figure 3) for different PyramidApp stages, to avoid the problem of different submission/rating times and drop-outs (inactive participants) causing the pyramid progression to freeze during option submission and rating stages. Submission timer and rating timer are the two main timers that define the maximum allowed time to complete those phases and their values can vary from minutes to days based on the activity being face-to-face or distance. To maintain a fluid dynamic flow, we use a satisfaction percentage (minimum number of active users completing a particular phase). Upon reaching the satisfaction percentage in a group, a countdown timer (*countdown timer < maximum allowed time per phase*) is activated until the maximum time allocated for that phase. The countdown timer notification is displayed in the interface for students to be informed. If all learners complete the task before any timer (either submission/rating timer or countdown timer) expiration, the group is promoted to the next level of the pyramid.

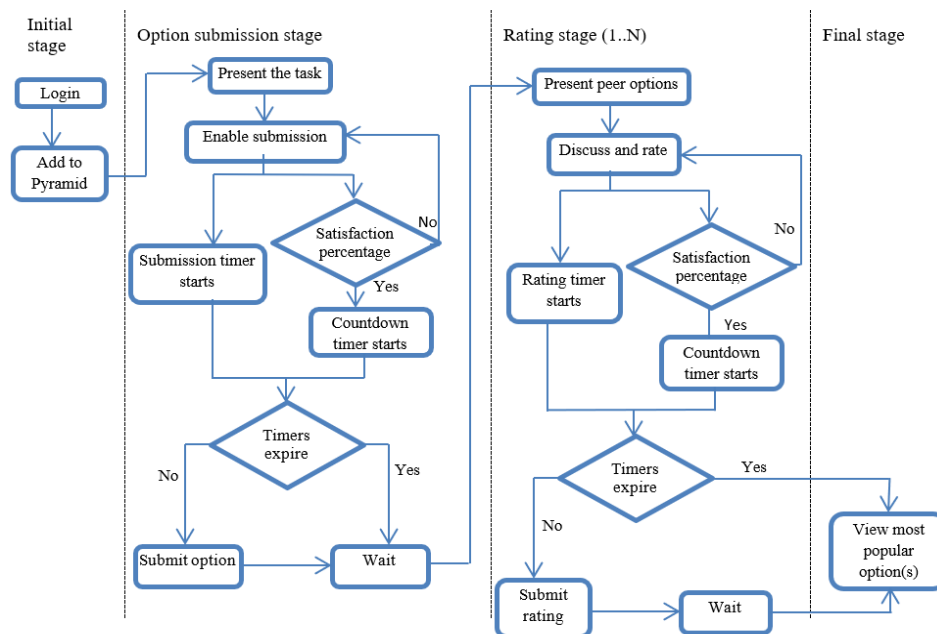


Figure 3. Flow control mechanisms showing dynamic flow progression.

Finally, we have integrated Pyramid flow activity awareness features that trigger information related to the activity status such as current pyramid level details, group members, countdown timer notifications and email notifications providing activity updates. Pyramid participants can see how many peers or groups who have not yet completed the current level during the waiting stage. Once the pyramid is finished, both practitioners and participants can view the highly rated option(s) resulted from the activity, which then could be -for example- further discussed and analyzed by the practitioner with the participants.

### Evaluation

Flexibility of the flow orchestration can be construed by means of elasticity and dynamism. Our working hypothesis is that the proposed flow creation and control mechanisms embedded to Pyramid CLFP address flexibility successfully. Hence, the evaluation questions articulated and analyzed in this study are, do proposed mechanisms for flow creation address elasticity and do proposed flow control mechanisms address dynamism with meaningful orchestration? Meaningful orchestration is being pedagogically relevant during the script execution, which means that any novel mechanism introduced to the flow does not violate the essence of the pattern: e.g. pyramid group sizes to be preserved necessarily, every participant should have at least one peer for

collaborations irrespective of activity drop-outs, late-comers are combined with on-going pyramids from the next possible activity phase and let them collaborate, propose default field values for the practitioners to create efficient Pyramid flows like possible number of pyramid levels or preferred group sizes and provide activity awareness measures for both participants and practitioners.

## Experimental settings and data gathering

Several rounds of experiments were carried out in several sessions from two undergraduate courses (Introduction to Information and Communication Technologies (ITIC) offered in the first year and Network Protocols (NP) offered in the third year) of an Engineering School, taught by the same professor. PyramidApp activity authoring requirements varied based on different epistemic tasks and target groups (see Table 1). Some experiments were enacted during face-to-face classroom sessions whereas others were in distance mode, asynchronously. Some PyramidApp activities were administered in a Massive Open Online Course (MOOC) called Innovative Collaborative Learning with ICT (CLAT), a five-week MOOC ended in summer 2017, launched on Canvas platform. To evaluate the proposed flow creation and control mechanisms, diverse attributes related to the mechanisms were introduced at activity authoring and enactment phases in PyramidApp. Various data sources such as activity log files and questionnaires were used as data collection methods. Log files provided more accurate data to analyze time durations, activity participation and behavior during PyramidApp activities. Questionnaires provided practitioner’s viewpoint as well as participants’ perspective towards the activity. We used a mixed approach for the analysis triangulating both quantitative and qualitative data gathered from above data sources to answer the research question addressed (Twining, Heller, Nussbaum, & Tsai, 2017). Quantitative figures helped to evaluate how successfully proposed parameters could implement elasticity and dynamism whereas qualitative data provided better interpretation for the results acquired in each scenario.

Table 1: PyramidApp experiment settings

| Activity Name                    | Activity Description  | Activity type  | Target group  | Additional Details   |
|----------------------------------|---|--|---|--|
| <b>ITIC (face-to-face class)</b> | Three cases were given to read. Initially the class was divided into halves and gave two cases to be discussed using PyramidApp. Then the final case was given to the whole class and enacted one PyramidApp round. | Case study analysis, open-ended question answering and collaborative negotiation | First year undergraduates (N=31)                      | Students used only one smartphone per small group of two or three students at the first rating level.                                      |
| <b>ITIC (distance mode)</b>      | Watch a video (discussing ethical dilemmas in ICT) and indicate which of the 24 imperatives in the ACM code of ethics and Professional Conduct are related.   | Case study analysis and collaborative negotiation                                | First year undergraduates (N=194)                     | Activity was enacted over a weekend as a homework before the next session using either smartphones or laptops individually.                |
| <b>NP (face-to-face class)</b>   | By observing the given TCP traffic, find some congestion control problems presented and explain your answer.  | Problem solving and collaborative negotiation                                    | Third year undergraduates (N=39)                      | Most students used smartphones individually. Activity was challenging. As expected by the educator, finally selected answer was incorrect. |
| <b>CLAT MOOC (distance mode)</b> | In your view, what are the main benefits of collaborative learning.   | Reflections upon practices and collaborative discourse                           | Educators from secondary and higher education (N=617) | Heterogenous user groups with diverse expertise levels and experiences used their own devices.   |

## Results and discussion

Scenarios stated in Table 1 have been analyzed indicating PyramidApp authoring configurations composed by the educators over resulted values during the activity enactment. Each case shows how the introduced flow creation and control mechanisms are used for pyramid formulation and how those achieved elasticity and dynamism. In terms of the meaningful orchestration achieved through the mechanisms, it was observed that there was no violation to the rules like maximum pyramid size or collaboration group sizes. Though some participants were dropped without completing all pyramid levels, participants had at least one peer to discuss during the collaboration stages and every group finally witnessed at least one solution irrespective of the number of initial submissions. Some participants joined the activity after the flow was initiated, yet pyramids occupied late-comers without interruptions. PyramidApp authoring features suggested default field values for the practitioners like possible number of levels in a pyramid or preferred group sizes and provided activity awareness measures such as current status of the activity, different groups, their members and responses for both participants and practitioners culminating a meaningful activity flow.

### ITIC face-to-face class scenario

In this case, the educator designed three separate Pyramid flow designs (see Table 2) and each design resulted in a single pyramid after the execution also, as the class size was relatively small. Though the three pyramids were initiated with the minimum size, those had grown to occupy more students ensuring the elasticity of the designs. The satisfaction percentage provided for the activity was 60% and the two countdown timers had been activated 24 times overall. Irrespective of the countdown timer expirations students were able to still submit the initial option and the ratings without pyramids being frozen ensuring fluid, dynamic pyramid progression until the submission phase timer expires. All three pyramids had consumed around 10 minutes for the activity completion, as desired by the practitioner.

Table 2: Flow creation and control mechanisms – ITIC in-class activity

| PyramidApp authoring parameters and values |                          |               |                              |                                    |                                    |                                  |                                  |  |                                       |
|--|--------------------------|---------------|------------------------------|------------------------------------|------------------------------------|----------------------------------|----------------------------------|--|---------------------------------------|
| Pyramid ID                                 | No. of pyramids designed | Class size    | Minimum students per pyramid | No. of levels                      | Group size (first level of rating) | Submit timer                     | Submit count down timer          | Rating timer                                     | Rating count down timer               |
| P1   | 1                        | 15            | 15                           | 2                                  | 3                                  | 4 mins                           | 2 mins                           | 4 mins   | 2 mins                                |
| P2   | 1                        | 7             | 7                            | 2                                  | 3                                  | 4 mins                           | 2 mins                           | 4 mins   | 2 mins                                |
| P3   | 1                        | 7             | 7                            | 2                                  | 3                                  | 4 mins                           | 2 mins                           | 4 mins   | 2 mins                                |
| After execution of the PyramidApp activity |                          |               |                              |                                    |                                    |                                  |                                  |  |                                       |
| Pyramid ID                                 | No. of pyramids created  | No. of logins | No. of options               | Submit count down timer expiration | Submit timer expiration            | No. of students rated in level 1 | No. of students rated in level 2 | Rating count down timer expiration (both levels) | Rating timer expiration (both levels) |
| P1   | 1                        | 18            | 15                           | 3                                  | 3                                  | 16                               | 15                               | 13   | 1                                     |
| P2   | 1                        | 12            | 6                            | 3                                  | 6                                  | 10                               | 10                               | 3  | 2                                     |
| P3   | 1                        | 10            | 4                            | 2                                  | 6                                  | 8                                | 8                                | 0  | 2                                     |

### NP face-to-face class scenario

This is also a face-to-face classroom scenario in which the educator designed only one pyramid, but the flow creation rules replicated the design and created two pyramids on-demand to accommodate all activity participants (see Table 3). Though the educator had expected 40 students for the activity, only 32 were present on that day and that miscalculation had no effect on the PyramidApp enactment due to the elasticity mechanisms applied by the application. Here also the satisfaction percentage was set to 60% and 15 submissions had been done after the countdown timer was initiated and 19 students had rated after seeing the rating countdown timer. The fact that every student not rating both levels had no effect in the pyramid progression due to the dynamism mechanisms proposed which provided flexible orchestration till the final level of pyramids. Both pyramids consumed around 9 minutes to complete all the levels including discussions. The task was authored with more time for submission and rating deliberately by the educator as the task was very challenging and wanted students to fail, to establish the conditions of a motivated and rich discussion in the classroom about why they failed, and which would be the right answer. Students enjoyed the activity irrespective of being failed to answer.

Table 3: Flow creation and control mechanisms –NP in-class activity

| PyramidApp authoring parameters and values |                          |               |                              |                                    |                                    |                                  |                                  |  |                                       |
|--|--------------------------|---------------|------------------------------|------------------------------------|------------------------------------|----------------------------------|----------------------------------|--|---------------------------------------|
| Pyramid ID                                 | No. of pyramids designed | Class size    | Minimum students per pyramid | No. of levels                      | Group size (first level of rating) | Submit timer                     | Submit count down timer          | Rating timer                                     | Rating count down timer               |
| P1   | 1                        | 40            | 16                           | 2                                  | 6                                  | 5 mins                           | 2 mins                           | 3 mins   | 1 min                                 |
| After execution of the PyramidApp activity |                          |               |                              |                                    |                                    |                                  |                                  |  |                                       |
| Pyramid ID                                 | No. of pyramids created  | No. of logins | No. of options               | Submit count down timer expiration | Submit timer expiration            | No. of students rated in level 1 | No. of students rated in level 2 | Rating count down timer expiration (both levels) | Rating timer expiration (both levels) |
| P1   | 2                        | 16            | 11                           | 9                                  | 3                                  | 12                               | 9                                | 11   | 2                                     |
| P2   |                          | 16            | 10                           | 6                                  | 4                                  | 12                               | 14                               | 8  | 2                                     |

### ITIC distance scenario

This distance mode of PyramidApp activity was enacted as a homework task over a weekend in a relatively large class (n=194). As given in Table 4, the educator designed a distance mode pyramid activity and assigned 16 as the minimum pyramid size. The activity resulted in 6 pyramids that have grown till the maximum volume successfully complying with proposed elasticity mechanisms. Longer timers were assigned for submission and rating since this activity extended over two days in the distance mode. As for the flow control mechanisms, this

version of distance mode PyramidApp did not implement the countdown timer based on the satisfaction percentage. Instead, it showed the remaining timer notification for each phase from the beginning of the activity and notified students via email notifications. Hence, the analysis considers only the submission and rating timer expiration. Still, that did not affect the flow of the pyramids. In all pyramids, number of students rated and collaborated in the second level is lesser than the first level. Around 48% could not complete the submission phase (in P2, P3, P4) because they were added to the on-going pyramids as they had accessed the activity late. Yet, these students were given chance to participate and present opinions in the rating stages assuring meaningful orchestration along the Pyramid flow, rather letting them to be idle till the next activity is available. If students login after pyramid creation timestamp (i.e., submission timer expired), they are straightaway added to the next available on-going pyramid, allowing them to collaborate in rating and discussion stages.

Table 4: Flow creation and control mechanisms – ITIC distance activity

| PyramidApp authoring parameters and values |                          |               |                              |                         |                                    |                                  |                                       |
|--|--------------------------|---------------|------------------------------|-------------------------|------------------------------------|----------------------------------|---------------------------------------|
| Pyramid ID                                 | No. of pyramids designed | Class size    | Minimum students per pyramid | No. of levels           | Group size (first level of rating) | Submit timer                     | Rating timer                          |
| P1   | 1                        | 120           | 16                           | 2                       | 5                                  | 18 hrs                           | 12 hrs                                |
| After execution of the PyramidApp activity |                          |               |                              |                         |                                    |                                  |                                       |
| Pyramid ID                                 | No. of pyramids created  | No. of logins | No. of options               | Submit timer expiration | No. of students rated in level 1   | No. of students rated in level 2 | Rating timer expiration (both levels) |
| P1   | 6                        | 31            | 10                           | 21                      | 23                                 | 16                               | 7                                     |
| P2   |                          | 31            | 16                           | 15                      | 27                                 | 19                               | 1                                     |
| P3   |                          | 31            | 16                           | 15                      | 23                                 | 16                               | 4                                     |
| P4   |                          | 31            | 16                           | 15                      | 27                                 | 18                               | 2                                     |
| P5   |                          | 31            | 25                           | 6                       | 21                                 | 18                               | 4                                     |
| P6   |                          | 31            | 20                           | 11                      | 13                                 | 11                               | 14                                    |

### CLAT MOOC scenario

In the MOOC scenario, the educator designed a Pyramid collaborative activity that extended to three pyramids at the end of the course (see Table 5). Only P1 had the minimum number of participants (four students) whereas P2 and P3 had occupied more, preserving the elasticity properties introduced. Here also the distance version of the PyramidApp including timers only for submission and rating along with the email notifications was used. Here also the timers were quite longer than in a usual face-to-face class. Around 67% from the participants had submitted their options and 60% had rated at least one level of the activity. Though, the activity participation is not equal among all students, the pyramids fluidly finished with no freezing in any branch.

Table 5: Flow creation and control mechanisms – CLAT MOOC distance activity

| PyramidApp authoring parameters and values |                          |               |                              |                         |                                    |                                  |                                       |
|--|--------------------------|---------------|------------------------------|-------------------------|------------------------------------|----------------------------------|---------------------------------------|
| Pyramid ID                                 | No. of pyramids designed | Class size    | Minimum students per pyramid | No. of levels           | Group size (first level of rating) | Submit timer                     | Rating timer                          |
| P1   | 1                        | 4             | 4                            | 2                       | 2                                  | 2 hrs                            | 2 hrs                                 |
| After execution of the PyramidApp activity |                          |               |                              |                         |                                    |                                  |                                       |
| Pyramid ID                                 | No. of pyramids created  | No. of logins | No. of options               | Submit timer expiration | No. of students rated in level 1   | No. of students rated in level 2 | Rating timer expiration (both levels) |
| P1   | 3                        | 4             | 3                            | 1                       | 1                                  | 2                                | 2                                     |
| P2   |                          | 6             | 3                            | 3                       | 2                                  | 3                                | 3                                     |
| P3   |                          | 5             | 4                            | 1                       | 4                                  | 1                                | 1                                     |

### Conclusion

This paper offers a technology-oriented contribution to the orchestration technology research line within CSCL, aiming at facilitating the application of collaborative learning scripts from small to large settings in a flexible way. We have defined script flexibility in terms of *elasticity* and *dynamism* of the collaborative learning activity flow. Moreover, we have proposed mechanisms to address them for the case of scripts structured according to the Pyramid or Snowball collaborative learning flow pattern. A set of flow creation mechanisms (e.g. allocation of participants to pyramid groups considering ranges in desired group sizes) has been defined to enable the elastic incorporation of a varying number of participants to a Pyramid flow in an automatic manner. Flow control mechanisms (including timers and satisfaction parameters) are proposed to maintain a fluid, dynamic pyramid flow execution. These mechanisms have been implemented in the PyramidApp tool. Validation of PyramidApp across different educational settings showed that the flow creation and control mechanisms

introduced to the Pyramid flow achieved elasticity and dynamism. Dynamically growing pyramids and replication of pyramids on-demand accommodated late-comers. Dropping out from current activity did not harm the pyramid progression. The mechanisms also led to ensure that orchestration (or real time script management) aspects were meaningful, i.e. in alignment with the pedagogical structural elements of the Pyramid pattern. For example, even if every participant did not submit an initial option (e.g. if they arrive late or leave) for the given task in above cases, the mechanisms ensured every group had at least two participants and one option to discuss in the first rating level and the flow was not interrupted. The same strategy has been used at other rating stages too, assuring that the pyramid can level-up. Future work should study to what extent these mechanisms can be extrapolated to other script families (e.g. those based on the Jigsaw flow pattern).

The results also suggest that human and intelligent agent interventions could further improve the utility of the proposed mechanisms. Usage of the in-built discussion board of the PyramidApp was not satisfactory across experiments. Email notifications used in the distance mode, notifying activity status did not catch sufficient participant attention. Hence, in future discussion prompts, cues or agent technologies like learning companions can be introduced to study how they could aid in promoting higher engagement in discussions. Future improvements like allowing small groups to modify or submit new options after collaborating in the rating levels and enhancing the notification system could be beneficial. Moreover, an extended version of the monitoring dashboard that would enable human intervention by the educator could improve the options for orchestration support, e.g. by facilitating the participation of the educator in discussions when especially needed (e.g. alerted by the system) or to modify timers to regulate the activity progression if required.

## References

- Davis, W.A. (2002). A comparison of pyramids versus brainstorming in a problem-based learning environment. In *Focusing on the student: Proceedings of 11<sup>th</sup> Annual Teaching Learning Forum*. Perth: ECU.
- Dillenbourg, P., & Tchounikine, P. (2007). Flexibility in macro-scripts for CSCL. *Journal of Computer Assisted Learning*, 23(1), 1-13.
- Dillenbourg, P., Sanna, J., & Fischer, F. (2009). The evolution of research on computer-supported collaborative learning. *Technology-Enhanced Learning*, 3–19. Netherlands: Springer.
- Dillenbourg, P., & Jermann, P. (2010). Technology for Classroom Orchestration. In M.S. Khine & I.M. Saleh (Eds.), *New Science of Learning: Cognition, Computers and Collaboration in Education*, 45–52. New York: Springer.
- Hernández-Leo, D., Villasclaras-Fernández, E. D., Asensio-Pérez, J. I., Dimitriadis, Y., Jorrín-Abellán, I. M., Ruiz-Requies, I., & Rubia-Avi, B. (2006). COLLAGE: A collaborative Learning Design editor based on patterns. *Educational Technology and Society*, 9(1), 58–71.
- Kobbe, L., Weinberger, A., Dillenbourg, P., Harrer, A., Hämäläinen, R., Häkkinen, P., & Fischer, F. (2007). Specifying computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning*, 2(2–3), 211–224.
- Manathunga, K., Hernández-Leo, D. (2017). Authoring and enactment of mobile pyramid-based collaborative learning activities. *British Journal of Educational Technology*, 49(2), 262-275.
- Pérez-Sanagustín, M., Burgos, J., Hernández-Leo, D., & Blat, J. (2011). CLFP intrinsic constraints-based group management of blended learning situations. *Studies in Computational Intelligence*, 350, 115–133.
- Rodríguez-Triana, M. J. (2014). *Linking scripting and monitoring in blended CSCL Linking scripting & monitoring support in blended CSCL scenarios*. Unpublished Doctoral Dissertation, University of Valladolid, Valladolid, Spain.
- Roschelle, J., & Teasley, S. D. (1995). The Construction of Shared Knowledge in Collaborative Problem Solving. In C. O'Malley (Eds.), *Computer Supported Collaborative Learning* (pp. 66-97). Heidelberg: Springer.
- Sharples, M. (2013). Shared orchestration within and beyond the classroom. *Computers and Education*, 69, 504–506.
- Twining, P., Heller, R., Nussbaum, M., & Tsai, C. (2017). Some guidance on conducting and reporting qualitative studies. *Computers and Education*, 106, A1–A19.

## Acknowledgements

DHL is a Serra Hünter fellow. This work has been partially funded by “la Caixa Foundation” (CoT project, 100010434) and the National Research Agency of the Spanish Ministry of Science, Innovations and Universities MDM-2015-0502, TIN2014-53199-C3-3-R, TIN2017-85179-C3-3-R.

# Please Don't Shoot the Messenger! Prompts in Online Learning Groups - Influences of Nudging Messages' Sender and Publicness on Recipients' Perception and Attribution

Filipa Stoyanova, University of Duisburg-Essen, filipa.stoyanova@uni-due.de  
Nicole C. Krämer, University of Duisburg-Essen, nicole.kraemer@uni-due.de

**Abstract:** Co- and socially shared regulation of learning is a challenge for groups in diverse environments – offline and online (Järvelä et al., 2014). Specifically, prompting others to do their share of the work, might be challenging for interpersonal relationships. Thus, addressing the social psychological dynamics in online learning environments, an online experiment was conducted ( $N = 352$ ). In a  $2 \times (2 \times 2)$  between-subjects design the nature of the sender (group member/system), the sender's past proficiency (high/low) and publicness (private/public message) were varied to explore whether automated mediation can reduce detrimental effects and enhance impartial message perception, sender impression and causal attribution. Contrasts by sender and proficiency demonstrate improved message perception, sender impression and internal causal attribution when prompts were sent by the system instead of average or low proficient fellow students. Among all groups, public nudges indicate more negative, but also more persuasive perception, however system public and private nudges did not differ.

**Keywords:** nudging, CRL, SSRL, automated group support, mediation perception, attribution

## Introduction

High dropout-rates, low levels of participation and course satisfaction are common problems of large online courses (Erdmann et al., 2017). One didactic element that is hailed as potential solution to both, low motivation as well as low understanding is small group cooperation within large online courses. Thereby, "... the interaction process is considered to be a more important element in learning than the outcomes" (Vuopala, Hyvönen, & Järvelä, 2016, p. 26). It has indeed been demonstrated that small group work is related to higher course satisfaction (Bernard et al., 2009). However, to be successful, small group work requires participation and is dependent on group development processes (Walther & Bunz, 2005). Especially when only consisting of few members, the development and existence of small groups might be endangered by dropout-rates and delays of group activities. In order to achieve beneficial social and educational effects, it has been shown to be advantageous when groups are supervised and supported by tutors (Thorpe, 2002). However, in large online courses (e.g. MOOCs) with a high number of small groups, time and staff related capacity is unaffordable. Therefore, one solution might be automated mediation which facilitates interaction in small-groups and prevents common unproductive behaviors. "Nagging" others to do their share of the work has already been identified by educational and social psychologists as one of eight common detrimental problems in small-group collaboration (Strauß, Rummel, Stoyanova, & Krämer, 2018) that might be addressed by an automatic tutor system.

## Theoretical background

So far, "nagging" or nudging other group members is barely researched as a social psychological challenge for interpersonal relationships in small group collaboration, hence it has not been clearly defined or conceptualized. The term "nudging" has been defined as a form of soft paternalism helping to defeat cognitive or behavioral biases in decision making, as "any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives" (Wang et al., 2013, p. 376). Nudging is the case, for instance, when displaying driver's current speed instead of forcing them to slow down. Such indirect means have been successfully applied in the context of health or food and for the purpose of online privacy (e.g. Acquisti et al., 2017). Nudging in the current study refers to hints in order to regulate others' activities, i.e. *unbalanced* co- or socially shared regulation of learning (Isohätälä, Järvenoja, & Järvelä, 2017). Thus, in the context of learning groups, nudges can be understood as negative feedback, co- or socially shared regulation of learning, that might lead to task and person conflicts and can be assumed as either beneficial or detrimental. For instance, effective non-individual regulation of learning, as well as task conflicts are beneficial due to the facilitation of group discussions, decision making, performance and knowledge building (Janssen, Van de Vliert, & Veenstra, 1999), but failures and their effects on interpersonal relationships are yet unknown. Furthermore, socially shared regulation of learning was shown not to function well in case of disengagement,

unequal participation and contribution (Isohätälä et al., 2017). Specifically person conflict reduces teams' effective communication and cooperation, tolerance for opposition, and openness to ideas of disliked team members, and simultaneously improves hostile attributions of others' intentions and behaviors (Janssen et al., 1999). Relationship conflict also was found to be linked to decreased perception of task performance (Mohammed & Angell, 2004). While under certain circumstances the more beneficial and desirable task conflict can facilitate group discussions, these can easily transform into more detrimental person conflicts (e.g., due to the fact that group members can hardly distinguish between them, Janssen et al., 1999). An explanation for the misinterpretation of task conflicts as person conflicts can be the need for consistency, i.e. disliked positions regarding the task can lead to corresponding personal attitudes toward the author. Furthermore, even justified criticism can generate reciprocal disliking and conflicts (Ilgen, Mitschell, & Frederickson, 1981).

Negative feedback has also been shown to have detrimental effects in groups (Gabelica, Van den Bossche, Segers, & Gijsselaers, 2012). Additionally, personal remarks and conflict hinder the success of groups' decision making process (Walther, 1996). Socio-emotional concerns like conflicts distract from the substantive task, and are especially time consuming in computer mediated communication (CMC), but can also easily get restricted on purpose in CMC settings (reducing channels, synchronicity etc.). Given the fact that nudging in learning groups occurs specifically based on missing contribution, it resembles feedback which has also been discussed as a possible challenge for interpersonal relationships. Feedback has been defined repeatedly depending on focus and context, as a comparison of actual performance and a desired standard (e.g. Gabelica et al., 2012). Beyond the aforementioned social-psychological perspective on group interaction as needed but potentially dysfunctional, peer interaction has been found to improve performance and learning outcomes – especially when guiding and support is present to avoid less efficient learning behaviors (Diziol, Walker, Rummel, & Koedinger, 2010). Feedback could be a supportive guiding solution, however, besides many benefits, it also turns out to be demotivating and leads to performance decreases (Kluger & DeNisi, 1996). Especially negative feedback can get unfavorable and result in inefficient behaviors (Ilgen & Davis, 2000). Conversely to the readily accepted positive feedback, negative feedback has been shown to be rejected and perceived as invalid and inaccurate (Ditto & Boardman, 1995). On the other hand, it has been found that compared to positive feedback, students prefer more negative feedback, as especially constructive if it points out a lacking goal progress instead of a decrease in engagement (for review see Fishbach, Eyal & Finkelstein, 2010). Similarly, negative feedback was concluded to have potential detrimental effects on members' affective reactions but may also have positive motivational effects (Gabelica et al., 2012), for instance regarding regulation of learning. Thus, research showed contradicting results and similar to conflicts, both detrimental and beneficial effects of negative feedback can be assumed.

Based on these findings it can be assumed that when one group member nudges another detrimental effects occur that endanger group climate and interpersonal relationships. Especially, once negative attribution of a certain group member or group conflicts occur, it can be more difficult to resolve conflicts and repair one's image or the wellbeing on a group level. One potential solution can be an external envoy as a mediator of "negative feedback" that would otherwise needed to be delivered by other group members.

## Relevance of feedback source

Feedback can be based on a subjective opinion or an objective measure and derive from an internal or external source. Here, group members might be categorized as an internal feedback source as well as providing a subjective opinion. Automated mediated feedback, such as messages deriving from the system as an external source, might be categorized as potentially based on more objective measures. Feedback is more likely perceived as accurate, when deriving from more credible, powerful or knowledgeable sources (e.g. London & Sessa, 2006). This, however, was mainly shown in the classic teaching context regarding learning processes and outcomes (e.g. Finn et al., 2009).

Gabelica et al. (2012) conclude that feedback intervention effectiveness might be improved if feedback is accurate, given in a timely manner, regular, given directly to teams it targets, *shared, non-threatening, and when its distribution is fairly equal*. However, these characteristics could vary depending on the source, as notably the last three could not be secured or held constant when feedback is given in a subjective manner by group members, but they can potentially be adjusted explicitly when providing feedback by an external system, as described in the next section regarding the media equation theory.

## Past proficiency

Regarding the effects of feedback, it is not only important who the sender is but also his/her attributes and prior behavior will play a role. One important aspect is past proficiency of group members as it has been described in group dynamic theories referring to the term of idiosyncrasy (i.e., possessing unique characteristics or showing unique behavior). In order to be allowed to derogate from a group's normative standards the individual member



must have shown high achievement in the past, i.e. must have earned a high level of “idiosyncrasy credit” (Hollander, 1958). In its origin, it considers leaders acquiring credit over time by performing continuously well and following group norms. As a credit of trust it enables one to deviate from the norms. Applied to learning groups this credit may be a basic factor in the case of the common problems of unequal contribution, social loafers and free riders.

As a recipient of a nudging message one could perceive it as inappropriate to receive any nudging remarks from a member with low past proficiency. Conversely, high past proficiency as a credit of trust and high contributor image as a kind of a group leader could turn into special rights to announce feedback within the group. This may be a fundament for the effectiveness of the message e.g. persuading the recipient to reconsider an act due to higher internal causal attribution.

### Media equation theory

Automated support is not simply an accurate and financially effective method to support high number of groups in online learning settings. In the sense of a mediator it could be perceived more impartial and can potentially prevent the occurrence of interpersonal conflicts in the case of nudging group members to do their share of the work. In case negative feedback is being sent by the system it can be less frustrating. As postulated in the media equation theory, interactions with computers and media can be perceived as real life interactions (Reeves & Nass, 1996). People have been demonstrated to treat computers similar to people, e.g. they avoid to deliver directly negative feedback. However, research in the area of information communication technology and human-robot-interaction has already shown some differences and limitations of the media equation theory. For instance, people asked to abuse a robot in an adaption of the Milgram experiment, more likely did so, when abusing a robot instead of a human (Bartneck, Rosalia, Menges, & Deckers, 2005). Furthermore, compared to other humans, virtual humans were shown to increase the willingness to disclose confidential information with them, although they are *only* computers (Lucas, Gratch, King, & Morency, 2014). Hence, it can be assumed that negative feedback from an automated system might be perceived differentially, e.g., as more impartial, since message recipients could not blame the system and its *personal* reasons for presenting feedback. Recipients could still indeed perceive the system as another human and interact with it equally, as the media equation theory postulates. However, the knowledge that “the system” has no past proficiency (neither high nor low) or any personal intentions can be significant for different perception. We assume that even if people blame the system, it would be on a level according to the missing, i.e. neutral past proficiency instead of high or low as in the case of other members’ feedback.

### Public versus private nudges

Beyond source, the way the message is presented needs to be scrutinized. One important aspect of presenting nudges in groups is whether these are targeted privately at the deviating person or displayed publicly in the group. The psychologically relevant construct in this context is ego threat. Negative feedback might pose an ego threat to one’s self-image or public-image (Audia & Locke, 2003). An ego threat of self-image provides information contrary to the own beliefs about the self, whereas for public-image respectively information contrary to one’s self targeted impression to others. Less or late contribution can be reasoned diversely, consciously or unconsciously, i.e. fitting or contrary to the self-image. Respectively nudges to hurry up and contribute more can be contrary to or fit the self-image too. However, since people generally tend to have a positive self-concept, prompting messages about their missing contribution could serve as a self-image threat. Indeed, threats to the self-image have already been operationalized by providing negative feedback (Leary, Terry, Allen, & Tate, 2009). Regarding the public-private distinction, the mere presence of others during feedback might be sufficient for a public-image threat since self-presentation is a basic need and ubiquitous. In the current study, it can emerge as soon as negative feedback is given in public in the group forum, i.e. others will also be informed about the presumable failure. Prior research on public and private feedback recommended it to be given in both ways in order to activate all potentially positive effects (Alvero, Bucklin, & Austin, 2001), but differences between both types may occur regarding nudging.

### Present study and hypotheses

Based on the reported theories and findings we aim to primarily explore which factors influence the perception of nudging and secondary, whether automated mediation (e.g. prompts sent by the system) can reduce potentially detrimental effects of nudging for interpersonal relationships and group climate. In an online study, in an artificial learning environment, prompt messages from the system as well as from team members will be conceptualized. Participants will imaginarily collaborate in small groups and receive prompting messages by means of visual

mock-ups. The nudging sender (group member vs. system), his proficiency (high vs. low engaged) and the publicness level (nudge in a private vs. group forum message) will be varied as independent variables.

Based on the literature summarized above, we assume differences between the experimental groups regarding causal attribution, emotional affect, sender impression and message perception. The media equation theory postulates, that people interact with machines like they do with humans. However, limitations of the theory already demonstrated different, partially more confidential interactions with machines. Hence, we assume that system nudges will be perceived differently than nudges from human team members – independent of proficiency and publicness. *H1: Compared to prompts from a team member, system prompts improve a) emotional affect, b) sender impression, c) message perception and d) internal causal attribution.*

Beyond this assumption, it cannot be derived whether interaction with the system is still beneficial, when contrasted to humans who showed relevant prior engagement and therefore earned idiosyncrasy credit. We therefore pose the following research questions: *RQ1.1: Compared to prompts from a low proficient team member, do system prompts improve a) emotional affect, b) sender impression, c) message perception and d) internal causal attribution? RQ1.2: Compared to prompts from a high proficient team member, do system prompts improve a) emotional affect, b) sender impression, c) message perception and d) internal causal attribution?*

Concerning the publicness of negative feedback and public ego-threat as hostile acts, we state that publicness has an impact on the negative emotional affect, on sender impression and message perception among all experimental groups: *H2: Publicness has an effect among all experimental groups, that increases a) negative emotional affect and b) negative message perception, but decreases c) persuasive message perception and d) sender impression.*

However, the main effect among all groups does not deliver details to scrutinize which publicness level of system nudges is perceived less threatening. Therefore we explore the effect of publicness in system treatment groups, exclude further influences, and state the research question: *RQ2: Among system treatment conditions, is there an impact of publicness on a) negative emotional affect, b) negative message perception, c) persuasive message perception and d) sender impression?*

## Method

Participants were instructed to imagine that they participate in an online learning group, did not provide their contribution shortly before the deadline and therefore received a prompting message. The nudging sender (group member vs. system), past proficiency of sender (high vs. low engaged) and the publicness level (nudge in a private vs. group forum message) were varied as independent variables. The design is not fully crossed as proficiency of the sender can only be varied in the team member conditions, not in the system conditions. To immerse the participants we created visual vignettes based on the view of a group member account in the learning environment Moodle and adapted them to the specific conditions (Figure 1). The group context was described in advance by short texts and charts on the past progress of the group and group members' past proficiency regarding contributions' quantity and timeliness. Participants were presented a prompt message in form of a visual mock-up either from the system or from a high or low proficient teammate, and either as a private (inbox) or public (forum) message. The use of an artificial learning environment allowed us to remind participants of the group context in the main message with the aid of a Moodle tool presenting group members' past online activities.

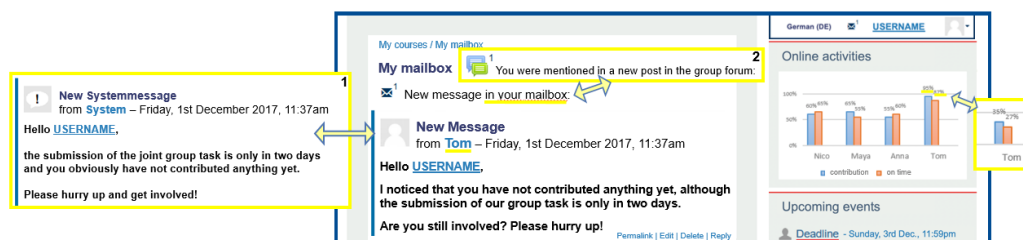


Figure 1. Exemplary vignette with variable conditions (1 = sender, 2 = publicness, 3 = proficiency).

Regarding the *sample*, the study was approved by the ethics committee of the University. A total of 444 participants were randomly assigned to one of the six conditions. They were mainly recruited through advertisements in Facebook groups and incentivized in a lottery. Additionally we used a crowdsourcing website with postpaid incentives. 92 persons were excluded from further analyses as they spent less than 10 seconds at the stimulus material pages (vignettes), their (nick)names were shorter than three letters and due to missing data. The remaining 352 participants (235 female, 117 male (33.2%)) ranged in age from 18 to 69 years ( $M = 29.40$ ,  $SD = 10.46$ ). Most of the participants had a university entrance degree (40.9%) or a higher degree (46.1%) and

were predominantly students (228, 64.8%) with a medium attitude towards group work participation ( $M = 2.95$ ,  $SD = 1.04$ , 1-5).

Regarding *measures*, adjective item lists were adapted from various affective scales and lists for emotional affect, message perception and sender impression. Finally, for each measure, an explorative factor analysis according to Horn (1965) was conducted and the recommended factor solution was chosen.

*Emotional Affect* was measured with a list of 23 adjective items, all employing a 5-point Likert scale (1 to 5 = strongly agree). Factor analysis revealed a 3-factor solution: *Positive* affect ( $\alpha = .841$ , 8 items, e.g. “inspired”), *negative external* affect ( $\alpha = .907$ , 10 items, “humiliated”), and *negative internal* affect ( $\alpha = .786$ , 5 items, “guilty”).

*Message Perception* was measured by 18 single items, employing a 5-point Likert scale (1 to 5 = strongly agree) and divided in 3 factors revealed by factor analyses: *Negative* ( $\alpha = .841$ , 5 items, “hostile”), *positive* ( $\alpha = .894$ , 7 items, “needed”) and *fair* ( $\alpha = .827$ , 6 items, “impartial”). Additionally, as a further perception dimension, persuasiveness of the message was measured with an adapted version of the perceived persuasiveness scale from Orji, Vassileva and Mandryk (2014), employing a 7-point Likert scale (1 to 7 = strongly agree). An additional self-generated item regarding reluctant behavior was included. One-factor solution was applied – *persuasive* ( $\alpha = .840$ , 5 items, “The prompt would persuade me”).

*Positive Sender Impression*. Semantic differentials were applied (e.g. 1 = unfriendly to 5 = friendly) to measure impression from the sender and taken according to factorial analysis as one-factor solution ( $\alpha = .905$ , 9 items).

*Causal Attribution*. We measured how participants attribute why they received the nudging message, whether it was their fault (internal) or others’ (external). Based on Lefcourt (1981), 8 items with daily internal and external reasons were generated on a 5-point Likert scale (1 to 5 = strongly agree). After a factor analysis the dimensions were combined in an overall factor *Internal causal attribution* ( $\alpha = .726$ , 5 items, “Because I was lazy”).

*Other Measures*. We employed one item measures to assess socio demographics (e.g. age, education, gender), past group work experiences (both quantity and valence), and attitude towards group work. We additionally measured self-esteem, perfectionism, and causal attribution style as traits, as well as further personality traits, which are not relevant for the analyses presented here.

## Results

Regarding *hypothesis 1*, after inspecting descriptive values (table 1), a planned comparison was conducted to test whether the system as a nudging agent compared to team members improves a) emotional affect, b) sender impression, c) message perception and d) internal causal attribution. Therefore, we compared both system-

Table 1: Descriptive values of dependent variables across treatment groups

| ↓DV / Treatment Group <sup>3</sup> | 1. SYpr |      | 2. SYpu |      | 3. TMpr- |      | 4. TMpr+ |      | 5. TMpu- |      | 6. TMpu+ |      | Total |      |
|------------------------------------|---------|------|---------|------|----------|------|----------|------|----------|------|----------|------|-------|------|
|                                    | M       | SD   | M       | SD   | M        | SD   | M        | SD   | M        | SD   | M        | SD   | M     | SD   |
| <i>Emotional Affect</i>            |         |      |         |      |          |      |          |      |          |      |          |      |       |      |
| Negative internal                  | 3.75    | .77  | 3.72    | .81  | 3.27     | .90  | 3.72     | .77  | 3.36     | .86  | 3.78     | .80  | 3.59  | .84  |
| Negative external                  | 2.47    | .96  | 2.66    | .82  | 2.57     | .80  | 2.62     | .92  | 2.82     | .94  | 2.88     | .97  | 2.67  | .91  |
| Positive                           | 2.22    | .73  | 2.34    | .75  | 2.05     | .75  | 2.32     | .76  | 2.16     | .72  | 2.23     | .79  | 2.22  | .75  |
| <i>Message Perception</i>          |         |      |         |      |          |      |          |      |          |      |          |      |       |      |
| Negative                           | 2.46    | 1.05 | 2.78    | .86  | 2.82     | .90  | 2.57     | .98  | 2.90     | .95  | 2.78     | .89  | 2.72  | .95  |
| Positive                           | 3.00    | .90  | 2.95    | .77  | 2.41     | .90  | 2.97     | .98  | 2.52     | .94  | 2.77     | .83  | 2.77  | .92  |
| Fair                               | 3.19    | .85  | 2.96    | .81  | 2.68     | .90  | 3.29     | .91  | 2.77     | .84  | 3.12     | .71  | 3.00  | .86  |
| Persuasive                         | 5.15    | 1.41 | 5.32    | 1.13 | 4.28     | 1.39 | 4.76     | 1.45 | 4.83     | 1.06 | 5.10     | 1.34 | 4.90  | 1.33 |
| <i>Positive Sender Impression</i>  |         |      |         |      |          |      |          |      |          |      |          |      |       |      |
|                                    | 3.38    | .66  | 3.21    | .69  | 2.86     | .85  | 3.35     | .80  | 2.84     | .90  | 3.23     | .70  | 3.14  | .80  |
| <i>Internal Causal Attribution</i> |         |      |         |      |          |      |          |      |          |      |          |      |       |      |
|                                    | 3.86    | .55  | 3.88    | .63  | 3.34     | .77  | 3.68     | .72  | 3.41     | .66  | 3.85     | .57  | 3.66  | .69  |

message-conditions (groups 1 & 2) in planned contrasts to all the team-member-conditions (groups 3, 4, 5 & 6). The contrast revealed significant differences, indicating increased levels in the system conditions regarding internal negative emotional affect,  $t(346) = -2.20$ ,  $p = .029$ ,  $r = .12$ , but also message perception positive,  $t(346) = -2.99$ ,  $p = .003$ ,  $r = .16$ , and persuasive,  $t(346) = -3.35$ ,  $p = .001$ ,  $r = .18$ , generally positive sender impression,  $t(263.12) = -2.76$ ,  $p = .006$ ,  $r = .17$ , and internal causal attribution,  $t(346) = -4.03$ ,  $p < .001$ ,  $r = .13$ . *Hypothesis 1* is partially supported since the general comparison of all treatment groups divided by sender showed significant differences and an improvement of the positive and persuasive message perception, sender impression, as well as higher levels of internal causal attribution, whereas, contradicting to *hypothesis 1*, the internal negative emotional

affect was higher. However, the system as a sender, compared to the groups with a team member sender, did not reveal significant differences regarding fair and negative message perception.

*Research question 1.1* and *1.2* investigate whether compared to a message sent by a team member with high or low past proficiency, a message sent by the system improves a) emotional affect, b) sender impression, c) message perception and d) internal causal attribution, we conducted planned contrasts. Due to the factor proficiency which was varied in all experimental groups with a team member as a sender, we had to split the analyses to avoid conflicting high and low proficiency to be summarized. Thus, we compared the system-message-conditions (1 & 2) in planned contrasts to the team-member-conditions divided by proficiency. For *RQ1.1* we compared them to those with low past proficiency (exp. groups 3 & 5). Contrasts revealed that a system message, compared to one from a low proficient team member increased internal negative emotional affect,  $t(346) = 3.97$ ,  $p < .001$ ,  $r = .21$ , message perception, (fair,  $t(346) = 3.19$ ,  $p = .002$ ,  $r = .17$ , positive,  $t(346) = 4.38$ ,  $p < .001$ ,  $r = .23$ , and persuasive,  $t(346) = 4.04$ ,  $p < .001$ ,  $r = .21$ ), positive sender impression,  $t(223.46) = 4.41$ ,  $p < .001$ ,  $r = .23$ , as well as internal causal attribution,  $t(346) = 5.81$ ,  $p < .001$ ,  $r = .22$ . There were no significant effects on external negative emotional affect and negative perception of the message. Except for these two subscales, the system-message-condition, compared to low proficient team-member-conditions, had an improving effect. For *RQ1.2* we compared both system-message-conditions (1 & 2) to those with high past proficiency team-member-conditions (exp. groups 4 & 6). However, there was no significant difference compared to high proficient team members. *Research question 1.1* revealed that messages from the system compared to those from low proficient team members improved internal causal attribution, positive sender impression, as well as fair, positive and persuasive message perception. However, on the other hand the system had detrimental effects regarding increased internal negative emotional affect. *Research question 1.2* did not reveal significant differences between the system and high proficient members.

*Hypothesis 2* was tested in a MANOVA, conducted among all treatment groups to test the influence of publicness on the experimental groups regarding increasing a) negative emotional affect and b) negative message perception, but decreasing c) persuasive message perception and d) sender impression. There was a significant effect of publicness,  $V = 0.64$ ,  $F(5, 346) = 4.77$ ,  $p < .000$ . Further separate univariate ANOVAs revealed significant effects on a) external negative emotional affect,  $F(1, 350) = 5.69$ ,  $p = .018$ ,  $\eta^2 = .016$ , b) negative message perception,  $F(1, 350) = 4.44$ ,  $p = .036$ ,  $\eta^2 = .013$  and c) persuasive message perception,  $F(1, 350) = 5.66$ ,  $p = .018$ ,  $\eta^2 = .016$ . However, there was no significant effect on d) sender impression. All significant effects indicated higher levels in public, rather than in private nudges, i.e. more negative emotional affect and more negative, but also more persuasive message perception. Therefore *H2* is partly supported as in the public condition a) negative emotional affect, b) negative message perception and c) persuasive message perception increased, but there was no significant effect on d) sender impression.

*Research Question 2* addressed the publicness of a system message and whether public and private system prompts differ regarding a) negative emotional affect, b) sender impression and c) negative and persuasive message perception. Therefore a MANOVA was conducted, by comparing solely the experimental conditions with a system sender. No significant differences were revealed,  $V = 0.12$ ,  $F(15, 99) = .893$ ,  $p = .574$ .

## Discussion

In order to gain the benefits of group conflict without the costs, we focus on the common, but barely researched nudging in groups. In an online experiment we addressed the social psychological dynamics in online learning groups and explored potentially influential factors for the perception of nudging, i.e. the sender, sender's past proficiency, as well as publicness of the message. The data indicated that nudging messages sent not by a team member but by the system, were perceived more positive and persuasive, improved sender impression and internal causal attribution. In line with prior research (Lucas, Gratch, King, & Morency, 2014) this indicates that a message from a human is not always equal to a message from the system and that it can be beneficial if a system instead of a fellow human delivers unpleasant messages. Contradicting our assumptions, however, internal negative emotional affect was also higher, potentially due to the fact that participants were more likely to blame themselves when the system confronted them with negative feedback. Also against our assumptions, the system messages were not perceived more impartial. Future research will need to show whether this – in the sense of the media equation (Reeves & Nass, 1996) – actually indicates that humans and machines are subject to the same person perception mechanisms or whether this result is due to different expectations towards machines and humans.

While there was no difference between highly proficient team members and the system, the system was perceived more positive compared to low proficient team members. Keeping the idiosyncrasy credit and the media equation theory in mind, participants may think of the system differently and still attribute a level of past proficiency to the system or at least attribute sufficient competences to the system to accept that it judges oneself. Alternatively, the system might have been accepted as a truly neutral evaluator who is allowed to utter feedback

just as much as a proficient peer is. In order to address these open questions, systems' error rates should be taken into account in future studies demonstrating the systems' past proficiency. Our findings further affirm some studies in educational artificial intelligence and tutoring systems. Compared to human-tutors, intelligent computer-tutoring was shown as equally and more effective independent of time and context (for review Kulik & Fletcher, 2016) and concluded as needed only if beneficial for performance and learning (Ostrander et al., 2019). However, the mere system perception and social-psychological group dynamics were disregarded.

Regarding limitations this study so far only elucidates the effect of system nudging in the artificial context of imaginary groups. Future field studies are needed in field setting and real groups to replicate the findings. A survey analysis was applied that primarily focused on perception, but behavioral data may be more promising to investigate the topic and its links to students' learning processes and outcomes. Finally, it has to be noted that all effect sizes were small.

In conclusion, having the tutor-system deliver nudges seems to be a promising solution for a specific form of group conflicts. Future research should also include and consider the system's potential embodiment and other cues such as natural language output. The more we learn about the conditions under which negative feedback can unfold positive effects (and the messenger does not have to be shot), the better can tutoring systems be improved to support group dynamics.

## References

- Acquisti, A., Adjerid, I., Balebako, R., Brandimarte, L., Cranor, L. F., Komanduri, S., . . . Wang, Y. (2017). Nudges for privacy and security: Understanding and assisting users' choices online. *ACM Computing Surveys (CSUR)*, *50*(3), 1–41. <http://dx.doi.org/10.1145/3054926>
- Alvero, A. M., Bucklin, B. R., & Austin, J. (2001). An objective review of the effectiveness and essential characteristics of performance feedback in organizational settings (1985–1998). *Journal of Organizational Behavior Management*, *21*(1), 3–29. [https://doi.org/10.1300/J075v21n01\\_02](https://doi.org/10.1300/J075v21n01_02)
- Audia, P. G., & Locke, E. A. (2003). Benefiting from negative feedback. *Human Resource Management Review*, *13*(4), 631–646. <https://doi.org/10.1016/j.hrmr.2003.11.006>
- Bartneck, C., Rosalia, C., Menges, R., & Deckers, I. (2005). Robot abuse – a limitation of the media equation. In *Proceedings of the Interact 2005 Workshop on Abuse*. Rome, IT.
- Bernard, R. M., Abrami, P. C., Borokhovski, E., Wade, C. A., Tamim, R. M., Surkes, M. A., & Bethel, E. C. (2009). A meta-analysis of three types of interaction treatments in distance education. *Review of Educational Research*, *79*(3), 1243–1289. <https://doi.org/10.3102/0034654309333844>
- Ditto, P. H., & Boardman, A. F. (1995). Perceived accuracy of favorable and unfavorable psychological feedback. *Basic and Applied Social Psychology*, *16*(1–2), 137–157. <https://doi.org/10.1080/01973533.1995.9646106>
- Diziol, D., Walker, E., Rummel, N., & Koedinger, K. R. (2010). Using intelligent tutor technology to implement adaptive support for student collaboration. *Educational Psychology Review*, *22*(1), 89–102. <https://doi.org/10.1007/s10648-009-9116-9>
- Erdmann, J., Rummel, N., Christmann, N., Elson, M., Hecking, T., Herrmann, T., . . . Wichmann, A. (2017). Challenges in implementing small group collaboration in large online courses. In B. K. Smith, B. K., M. Borge, E. Mercier, & K. Y. Lim (Eds.), *Making a difference: Prioritizing equity and access in CSCL, 12th International Conference on Computer Supported Collaborative Learning (CSCL) 2017*, Volume 2. Philadelphia, PA: International Society of the Learning Sciences.
- Fishbach, A., Eyal, T., & Finkelstein, S. R. (2010). How positive and negative feedback motivate goal pursuit. *Social and Personality Psychology Compass*, *4*(8), 517–530. <https://doi.org/10.1111/j.1751-9004.2010.00285.x>
- Finn, A. N., Schrodt, P., Witt, P. L., Elledge, N., Jernberg, K. A., & Larson, L. M. (2009). A meta-analytical review of teacher credibility and its associations with teacher behaviors and student outcomes. *Communication Education*, *58*(4), 516–537. <https://doi.org/10.1080/03634520903131154>
- Gabelica, C., Van den Bossche, P., Segers, M., & Gijssels, W. (2012). Feedback, a powerful lever in teams: A review. *Educational Research Review*, *7*(2), 123–144. <https://doi.org/10.1016/j.edurev.2011.11.003>
- Hollander, E. P. (1958). Conformity, status, and idiosyncrasy credit. *Psychological Review*, *65*(2), 117–127. <http://dx.doi.org/10.1037/h0042501>
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185. <https://doi.org/10.1007/BF02289447>
- Ilgen, D., & Davis, C. (2000). Bearing bad news: Reactions to negative performance feedback. *Applied Psychology*, *49*(3), 550–565. <https://doi.org/10.1111/1464-0597.00031>

- Ilgen, D. R., Mitchell, T. R., & Fredrickson, J. W. (1981). Poor performers: Supervisors' and subordinates' responses. *Organizational Behavior and Human Performance*, 27(3), 386–410. [http://dx.doi.org/10.1016/0030-5073\(81\)90030-1](http://dx.doi.org/10.1016/0030-5073(81)90030-1)
- Isöhätälä, J., Järvenoja, H., & Järvelä, S. (2017). Socially shared regulation of learning and participation in social interaction in collaborative learning. *International Journal of Educational Research*, 81, 11–24. <https://doi.org/10.1016/j.ijer.2016.10.006>
- Janssen, O., Van de Vliert, E., & Veenstra, C. (1999). How task and person conflict shape the role of positive interdependence in management teams. *Journal of Management*, 25(2), 117–141. <https://doi.org/10.1177/014920639902500201>
- Järvelä, S., Kirschner, P. A., Panadero, E., Malmberg, J., Phielix, C., Jaspers, J., . . . Järvenoja, H. (2014). Enhancing socially shared regulation in collaborative learning groups: designing for CSCL regulation tools. *Educational Technology Research and Development*, 63(1), 125–142. <https://doi.org/10.1007/s11423-014-9358-1>
- Kluger, A. N., & DeNisi, A. (1996). Effects of feedback intervention on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254–284. <https://doi.org/10.1037//0033-2909.119.2.254>
- Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research*, 86(1), 42–78. <https://doi.org/10.3102/003465431558142>
- Leary, M. R., Terry, M. L., Allen, A. B., & Tate, E. B. (2009). The concept of ego threat in social and personality psychology: Is ego threat a viable scientific construct? *Personality and Social Psychology Review*, 13(3), 151–164. <https://doi.org/10.1177/1088868309342595>
- Lefcourt, H. M. (Ed.). (1981). *Research with the locus of control construct*. Orlando, FL: Academic Press.
- London, M., & Sessa, V. I. (2006). Group feedback for continuous learning. *Human Resource Development Review*, 5(3), 303–329. <https://doi.org/10.1177/1534484306290226>
- Lucas, G. M., Gratch, J., King, A., & Morency, L. P. (2014). It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior*, 37, 94–100. <https://doi.org/10.1016/j.chb.2014.04.043>
- Mohammed, S., & Angell, L. C. (2004). Surface- and deep-level diversity in workgroups: Examining the moderating effects of team orientation and team process on relationship conflict. *Journal of Organizational Behavior*, 25(8), 1015–1039. <https://doi.org/10.1002/job.293>
- Orji, R., Vassileva, J., & Mandryk, R. L. (2014). Modeling the efficacy of persuasive strategies for different gamer types in serious games for health. *User Modeling and User-Adapted Interaction*, 24(5), 453–498. <https://doi.org/10.1007/s11257-014-9149-8>
- Ostrander, A., Bonner, D., Walton, J., Slavina, A., Ouverson, K., Kohl, A., ... & Winer, E. (2019). Evaluation of an Intelligent Team Tutoring System for a Collaborative Two-Person Problem: Surveillance. *Computers in Human Behavior* (in press). <https://doi.org/10.1016/j.chb.2019.01.006>
- Reeves, B., & Nass, C. I. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Chicago, IL: Center for the Study of Language and Information; New York, NY: Cambridge University Press.
- Strauß, S., Rummel, N., Stoyanova, F., & Krämer, N. (2018). Developing a library of typical problems for collaborative learning in online courses. In J. Kay & R. Luckin (Eds.) (2018), *Rethinking Learning in the Digital Age: Making the Learning Sciences Count, 13th International Conference of the Learning Sciences (ICLS) 2018, Volume 2* (pp. 1045-1048). London, UK: International Society of the Learning Sciences.
- Thorpe, M. (2002). Rethinking Learner Support: The challenge of collaborative online learning. *Open Learning: The Journal of Open, Distance and e-Learning*, 17(2), 105–119.
- Vuopala, E., Hyvönen, P., & Järvelä, S. (2016). Interaction forms in successful collaborative learning in virtual learning environments. *Active Learning in Higher Education*, 17(1), 25–38. <https://doi.org/10.1177/1469787415616730>
- Walther, J. B. (1996). Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. *Communication Research*, 23(1), 3–43. <https://doi.org/10.1177/009365096023001001>
- Walther, J. B., & Bunz, U. (2005). The rules of virtual groups: Trust, liking, and performance in computer-mediated communication. *Journal of Communication*, 55(4), 828–846. <https://doi.org/10.1111/j.1460-2466.2005.tb03025.x>
- Wang, Y., Leon, P. G., Scott, K., Chen, X., Acquisti, A., & Cranor, L. F. (2013). Privacy nudges for social media: an exploratory Facebook study. In *Proceedings of the 22nd International Conference on World Wide Web* (pp. 763–770). New York, NY: ACM. <https://doi.org/10.1145/2487788.2488038>

# How Augmented Reality Affects Collaborative Learning of Physics: A Qualitative Analysis

Apittha Unahalekhaka, Iulian Radu, and Bertrand Schneider  
a.unahalekhaka@gmail.com, iulian\_radu@gse.harvard.edu, bertrand\_schneider@gse.harvard.edu  
Harvard University, Graduate School of Education

**Abstract:** Augmented reality (AR) is a powerful visualization tool to support learning of scientific concepts across learners of various ages. AR can make information otherwise invisible visible in the physical world in real-time. In this study, we are looking at a subset of data from a larger study (N=120), in which participant pairs interacted with an augmented sound producing speaker. We explored the learning behaviors in eight pairs of learners (N=16) who participated in an unstructured physics activity under two conditions: with or without AR. Comparing behaviors between the two experimental conditions, we found that AR affected learning in four different ways: participants in the AR condition (1) learned more about visual concepts (ex: magnetic field structures) but learned less about nonvisual content (ex: relationship between electricity and physical movement); (2) stopped exploring the system faster than NonAR participants; (3) used less aids in exploration and teaching; and (4) spent less time in teaching their collaborators. We discuss implications of those results for designing collaborative learning activities with augmented reality.

## Introduction and related work

Numerous learners find it challenging to master concepts where abstract topics cannot be explicitly seen and experienced in everyday life. For this reason, different aids have been used to make these concepts tangible. In physics education, tools such as visual representations (e.g., graphs, diagrams, icons) and physical manipulatives (e.g., blocks, magnets, compass) are frequently used as teaching aids (Zacharia & Olympiou, 2011; Savinainen et al, 2015; Suyatna et al, 2017). One issue with the usage of multiple representations is that learners must understand and integrate them (Kohl et al, 2007; Ainsworth 2008). For instance, Ainsworth (2008) has found that to use different representations effectively, learners must understand the relationship between the representation and the concept they are learning and understand what information the representation is carrying. This is especially problematic when different representations are presented in different places at different times (Bujak et al, 2013). Accordingly, using hand gestures (e.g., deictic and iconic) to connect multiple concepts together can reduce cognitive load and improve students' learning (Alibali et al, 2012; Alibali et al, 2013).

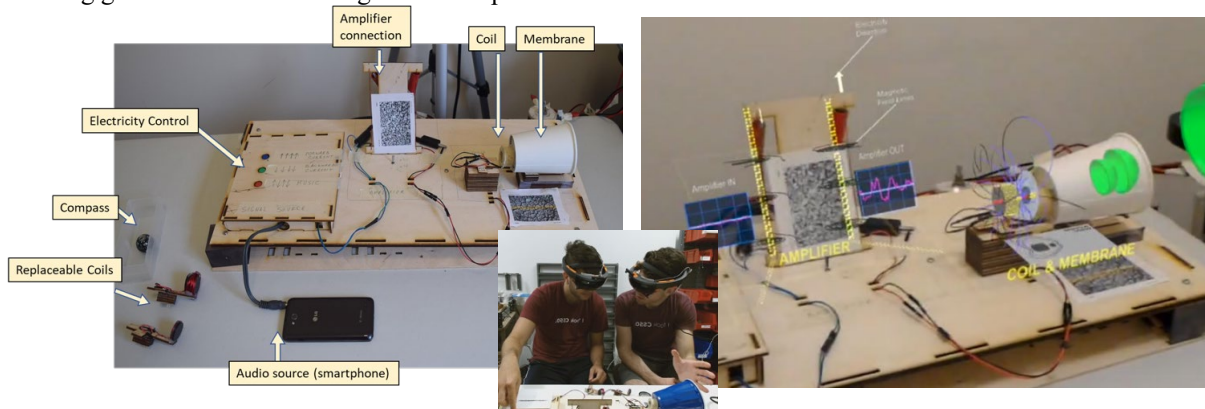
Augmented reality (AR) is an emerging technology that has become increasingly common to visualize abstract scientific learning (Wu, 2013). Augmented reality provides multiple affordances that can be beneficial for learning (Radu, 2014; Bujak et al, 2013). For example, AR can help reduce the cognitive load in learners while they are learning with multiple representations, by linking abstract representations to concrete physical representation through spatial and temporal contiguity (Radu, 2014; Bujak et al, 2013). This affordance is especially suitable for physics education, where abstract concepts accompany physical models. In one study of using AR for Newtonian force concepts, Enyedy et al (2012) found that children ages 6-8 years old were able to learn force concepts when AR was incorporated into role play. Cai et al (2017) revealed that AR-based motion-sensing tools could help junior high school students learn concepts about magnetic fields more intuitively, and that students were able to retain the concepts longer. Those findings suggest that AR can be an interesting medium for supporting collaborative learning. In the section below, we describe the larger experiment we ran and how we qualitatively analyze eight pairs from it.

## General description of the study

A larger study (Radu & Schneider, 2019) compared 60 pairs of participants (N=120) who interacted with an augmented physical model of an audio producing speaker (Figure 1). The pairs were assigned to either do the activity with or without seeing the augmented reality (AR) visuals. The NonAR condition participants were given a physical model, a compass, and informational posters on the wall (Figure 1, left), while the AR condition participants were given the same tools, and also various AR visuals in addition to what the NonAR participants were given (Figure 1, right). There were multiple representational tools to support participants in both conditions, with the main difference being that AR condition received visual representation aligned with the physical system, whereas the NonAR condition received representations not overlaid on the physical system.

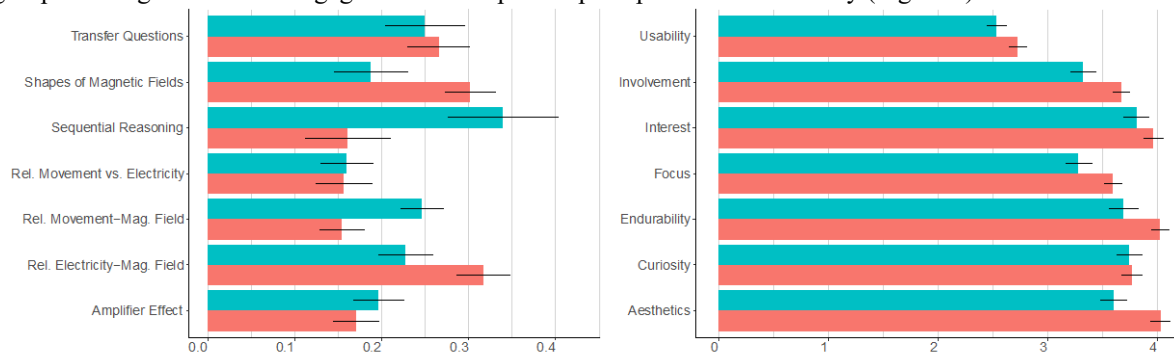


For this paper we focused on analyzing video recordings of eight pairs (n=16), which were chosen according to two variables: AR versus NonAR; and overall learning gains: high learning gains versus low learning gains. This is a 2x2 design with two pairs of students in each condition.



**Figure 1.** Physical model with no AR (left); users wearing two Microsoft Hololens® and interacting with the system (middle); Physical model with AR (right).

We found that participants in AR learned more about spatial structures (ex: shapes of magnetic fields) but learned less about non-visual concepts (ex: relationship between physical movement and electricity). Furthermore, AR groups had higher levels of engagement and improved perception of self-efficacy (Figure 2):



**Figure 2.** Group differences in relative learning gains in percentage (left); and overall attitudes in 5-point scale (right). Red= AR group, Green = NonAR group; whiskers show standard errors (for more information, please refer to (Radu & Schneider, 2019)).

In this paper we build upon this project by using the same dataset and qualitatively analyzing the learning and teaching behaviors of participants, comparing between AR-vs-NonAR conditions, by comparing groups that had high learning vs low learning. We are interested in understanding specifically why participants learned differently between the two conditions. Our three research questions are as follow:

Research Question 1: How does AR impact collaboration?

Research Question 2: How does AR impact the use of external aids and gestures?

Research Question 3: How does AR impact teaching behaviors?

## Method

This paper analyzes video recording of participants learning by using the system in the AR and NonAR conditions. During the activity two participants had to work together to complete a worksheet on electromagnetism; the worksheet asked questions such as “Where is the strongest magnetic field located?” and “Draw the shape of the magnetic field when a cup is closer versus further”. At fixed times during the experience, the facilitator would ask questions to participants, such as “How is the direction and the strength of the electric current influencing the cup?”, which were meant as provocation to think about various aspects of the system, but did not require explicit answering.

We performed qualitative video analysis to understand participant behavior. A coding scheme was constructed through iterative bottom-up coding. The final coding scheme consisted of three main categories: (1)



communication type; (2) aid provided for communication; and (3) method of communication. Research study videos from the eight participant sessions were split into 30 seconds time frames and assigned one or more codes. For inter-rater reliability, one rater coded 100% of the videos, while the other rater coded 20% of each video. The eight analyzed sessions ranged between 27-33 minutes. In total, 489 30-sec. segments were coded, accounting for 4 hours of video. Inter-rater reliability reached a Cohen kappa of 0.8, which implies substantial/almost perfect agreement. The coding scheme categories that were derived from the videos are as follows:

**Communication Types:** Describes the purpose of the participants' communication: *exploring the system*, *teaching each other*, *chatting about irrelevant topics*, or *non-interacting* (Table 1).

**Aid Provided for Communication:** If participants were using an aid while communicating, this code describes the type of tool or representational aid used: *poster*, *compass*, and *system*. The last category is coded separately for AR vs. NonAR groups because the system was merged with holographic representations for the AR group; this resulted in two codes for our coding scheme: *using NonAR system* (only applicable for NonAR participants, Figure 1 left), and *using AR system* (only applicable for AR participants, Figure 1, right).

**Methods of Communication.** If participants were using gestures or drawings as a method of communication, this code describes that method: *deictic gesture*, *iconic gesture*, and *self-drawing*. Deictic gesture is the type of method that is the easiest to produce, it is when a participant is pointing or using gesture to make the other participant shift his/her attention to a certain location (Roth, 2001). In contrary, iconic gesture and self-drawing has a higher representation level. Iconic gesture is a symbolic gesture when one participant is using his/her hand to mimic a visual, while self-drawing is when one participant is drawing on a paper to support his/her explanation.

Table 1: Definition and Examples of the Communication Types

| Communication Types | Definition   | Example  |
|---------------------|--|--|
| Exploring           | At least one of the two participants discuss about the activity with a clear intention to interaction with the other participant. Also includes non-verbal interaction when the participants test out the system together. | P1: "Alright, let's see what changes when you pull it down to 10x. Anything change?"<br>P2: " I think it's quite-er"<br>P1: "Yeah, I agree, it is definitely quite-er" |
| Teaching            | At least one of the two participants tries to make the other participant understands a physics concept through explanation or clarification.   | P1: "Look at the green lines over there. There are more lines inside and the magnetic field becomes bigger. The polarity is also changing north and south."            |
| Chatting irrelevant | Both participants talk on a topic unrelated to the physics concept based on the activity.  | P1: "Do you know a lot of physics?"<br>P2: "I forgot most of it"   |
| Non-interacting     | There is no explicit intention to interact between the two participants, but at least one participant is engaging with the learning activity.  | One participant played with the system, while the other sit still.   |

## Results

Results from the larger study (Radu & Schneider, 2019) indicate that there are differences in learning and attitudes across the AR vs NonAR conditions (Figure 2). When examining the eight pairs from this current study, we found similar results, namely that participants in the AR condition had higher learning gains in topics involving spatial structures (ex: identifying shapes of magnetic fields), while the NonAR condition participants had higher gains in topics related to physical movement (ex: relationship of magnetic field vs. movement). Additionally, similar to the overall results, participants in the AR condition had a higher tendency than NonAR participants to believe that physics is easy after completing the study. Figure 3 indicates that the eight groups we selected are representative of our sample.

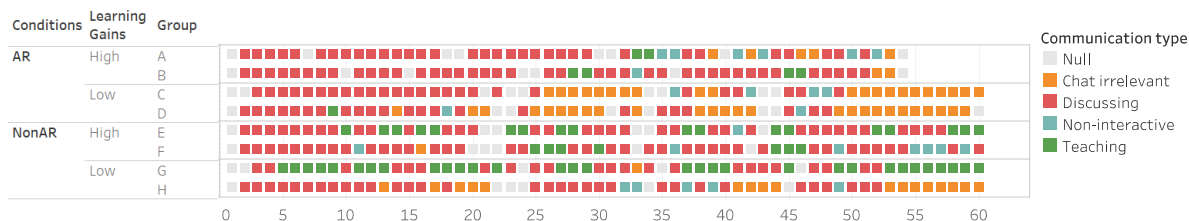
| Conditions | Learning Gains | Group | Curiosity | Physics is easy | Electricity vs. Magnetic Fields | Electricity vs. Movement | Magnetic Fields vs. Movement | Identifying Magnetic Field Shapes | Overall Total Relative Learning Gains |
|------------|----------------|-------|-----------|-----------------|---------------------------------|--------------------------|------------------------------|-----------------------------------|---------------------------------------|
| AR         | High           | A     | 0.17      | 0.25            | 0.49                            | 0.20                     | 0.35                         | 0.417                             | 0.44                                  |
|            |                | B     | 0.00      | 0.13            | 0.63                            | 0.00                     | 0.20                         | 0.583                             | 0.50                                  |
|            | Low            | C     | -0.17     | 0.13            | -0.33                           | 0.00                     | 0.20                         | 0.250                             | 0.10                                  |
|            |                | D     | 1.33      | 0.88            | 0.38                            | -0.03                    | 0.10                         | -0.225                            | 0.14                                  |
| NonAR      | High           | E     | -0.17     | -0.25           | 0.32                            | 0.33                     | 0.23                         | 0.300                             | 0.47                                  |
|            |                | F     | 0.50      | 0.25            | 0.10                            | 0.30                     | 0.40                         | 0.100                             | 0.42                                  |
|            | Low            | G     | -0.33     | -0.88           | 0.11                            | -0.17                    | 0.13                         | -0.100                            | 0.15                                  |
|            |                | H     | 0.50      | 0.00            | 0.18                            | 0.13                     | 0.45                         | -0.500                            | 0.18                                  |

**Figure 3.** Pre- and post- assessment: Difference in the attitude (curiosity and whether physics is easy) and learning scores on different areas (relationship: electricity-magnetic fields, electricity-movement, magnetic fields-movement; shapes of magnetic fields) between the 8 dyad groups.

### RQ1: How does AR impact collaboration?

We analyzed how participants across the two conditions spent their time during the study. Figure 4 shows the distribution of participant activities over the 30 minutes of the study. Participants in the AR condition spent only 73% of their session time actively engaged, with the remaining 27% of their time conversing on non-task-relevant topics. In contrast the NonAR condition spent 90% of time actively engaged, and 10% of time conversing on non-task-relevant topics. This pattern is more prominent in groups with low learning gains – low AR groups conversed about non-task relevant topics 47% of the time while low NonAR groups conversed about non-task topics 19%; the high AR groups spent 8% of their time on non-task relevant topics, vs. 1% in high NonAR. This indicates that AR groups have an earlier tendency to believe that they are finished the activity than compared to NonAR groups, and this effect was stronger in groups with low learning gains.

#### Communication



**Figure 4.** Time sequence for the communication behaviors in each 30 seconds block for the eight groups analyzed in this paper. Null=not exhibiting any of the listed communication types.

We illustrate the phenomenon of AR participants finishing the activity more quickly in Table 2. This example shows how differently AR and NonAR participants react to the same question asked by the facilitator. After hearing the question, the low AR group pause and stare at the AR system. Both participants observe the direction of the cup then come up with their assumption on the relationship between electricity-movement-magnetic fields. They appear to be satisfied with their answers and switch back to conversing on irrelevant topics. From this observation, this AR group seems to have a positive attitude towards the activity, but do not take effort to explore the questions more deeply. In contrast, it took longer for the NonAR group to discuss the same question. Participant 2 uses a lengthy explanation with various iconic hand gestures, while Participant 1 recaps. Subsequently, they use concepts from this discussion to extend their worksheet answer.

**Table 2: Quotes from participants in the Low AR group and Low NonAR group figuring out a question**

| Low AR Group   | Low NonAR Group  |
|--|--|
| Facilitator: “How is the direction and the strength of the electric current influencing the cup?”<br>[Facilitator leaves the room]<br>[P1 and P2 taking turns to press forward/backward buttons and look at the electromagnet with superimposed AR magnetic field]<br>P1: “Oh! It’s like the direction is either pushing it away or pulling it closer”<br>P2: “Yeah” | Facilitator: “How is the direction and the strength of the electric current influencing the cup?”<br>[Facilitator leaves the room]<br>P1: “Is the music linked to the current, so the stronger the current, the stronger the music is?”<br>P2: “Yeah, kinda, So the music is like, little tiny current signals like saying push pull, push pull, modulating. And then it gets amplified from pushing like this, to pushing |

P1: "The strength when pushing it away is less"  
 [P1 changing the magnitude of the amplifier and looking at the AR amplifier graph]  
 P2: "When the current is weaker, the impact on the membrane of the cup and the magnetic field is smaller"  
 P1: "Yeah"  
 P2: "I don't know where to add on the paper"  
 P1: "I think it's all there" [P1 pointing to their existing answers on the worksheet]  
 [Both sitting silently, then chatting on irrelevant topics]

like that" [P2 using iconic hand gesture showing different strength]. "The more windings you have, the stronger it gets. So you are shaking this magnet at a very specific pace, which then vibrates and makes sound" [P1 looking at P2 while explaining and making iconic hand gestures]  
 P1: "Do you mind answering the question, so the stronger the current, the greater amplification we hear from the cup?"  
 [Adding answer to the question that they previously skipped, while continuing to talk about the activity]

## RQ2. How does AR impact the use of external aids and gestures?

We examined how participants in the two conditions used aids and gestures to facilitate learning, by calculating the percentage of time each aid or gesture was used while participants were involved in either exploring or teaching. In the AR condition, the main learning aid was the system, used for the majority of the time (92%). In contrast, in the NonAR condition participants focused on the system 38% of the time, and spent more time using other aids (compass 32% and poster 12%). Figure 5 illustrates the sequence of switching between different aids. Additionally, the AR condition mostly used deictic gesture (47%) with little drawing (3%) and some iconic gesture (9%) to assist with their learning. This pattern contrasts from the NonAR condition that used less deictic gestures (34%), and much more drawing (14%) and more iconic gesture (12%). This suggests that some "tunnel vision" from the AR participants: they totally neglected other resources that were at their disposal.

### Aids

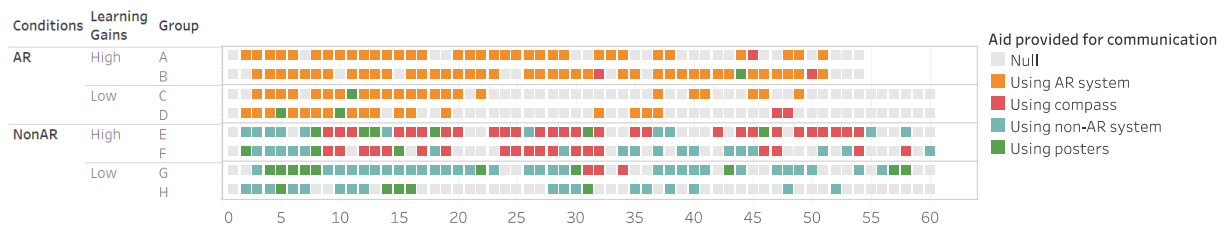


Figure 5. Time sequence for aids usage in each 30 seconds block. Null=not using any of the listed aids.

Figure 6 illustrates the sequence of gestures and drawings. These results indicate that the AR condition focused more on the AR-enhanced system and focused more on communication through pointing, while the NonAR participants used the system in concert with other external tools and communicated using iconic gestures and drawings. Again, this suggests more diverse behaviors from the NonAR participants – who used a variety of gestures and drawings to explore the concepts taught.

### Methods

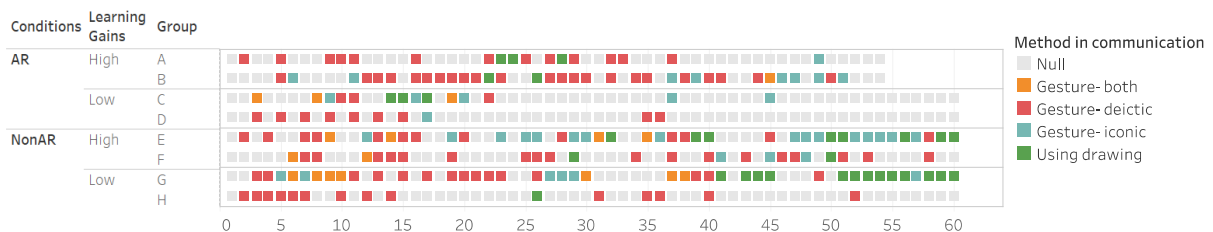


Figure 6. Time sequence for communication methods usage in each 30 seconds block. Null=not using any of the listed methods.

## RQ3: How does AR impact teaching behaviors?

While running the study, we observed differences in how participants explained concepts to each other. To further explore this question, we analyzed teaching behaviors between AR and NonAR groups. We found that in the AR condition, participants used 3.72% of the activity time for teaching each other, comparing to 27.39% in the NonAR

condition (Figure 7, left). Analyzing the use of aids in teaching, we found that AR participants only taught while using the system (100% of the teaching time). In the NonAR condition, participants taught using various aids: 33% of the time by using the system, 43% of the time using compass, 11% of the time using posters, and 12% using other methods such as drawings. Aside from the differences in systems/tools used as aids to teach, the participants in the two conditions also utilized different communication methods to teach (Figure 7, right): the AR condition participants mainly used deictic gestures (67% of the time), with little iconic gestures (8%) and no drawings (0%). In contrast, the NonAR condition participants used less deictic gestures (39%), and more iconic gesture (16%), and drawings (18%). These results suggest that the AR condition was able to teach more concisely by just using the AR system and skip the process of acquiring representations from other aids or communication methods needed by the NonAR groups.

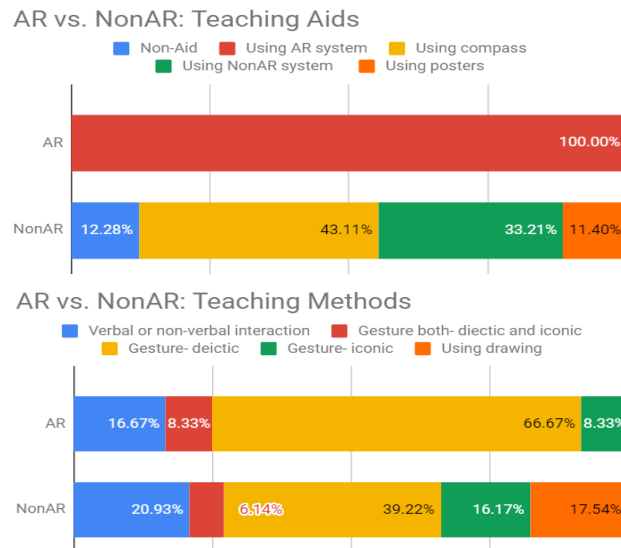


Figure 7. Percentage of time from the total teaching time spent on using each aids (left) and methods (right).

As an illustration of this difference, Table 3 shows how an abstract concept (magnetic polarity) is taught more easily with AR. Without AR, participants read polarity by using a compass, whose use required learning and experimentation. Participant 1 from the NonAR group (Table 3, right) slowly taught Participant 2 on how to read polarity from a compass, whereas in the AR group (Table 3, left) Participant 1 was able to instruct the other participant to look at the polarity labels (North / South) provided in AR near the AR magnetic fields. In this scenario, the AR group is able to skip the process of understanding how to physically read the magnetic polarity, and directly talked about alternative current. In contrast, the NonAR group took a longer time to learn about current in relation to magnetic direction, and remained focused on understanding the slow motions of the magnet rather than fast oscillations resulting in music.

Table 3: Quotes from participants in the AR group and NonAR group teaching

| AR Group  | NonAR Group   |
|---|---|
| <p>[Both looking at the magnetic field AR]<br/>           P1: "This button pushes the magnet in and out..."<br/>           P2: "Um hm"<br/>           P1: "Because the magnetic field changes, you see? It's inverted. It changes between north and south." [P1 referring to the AR magnetic field]<br/>           P2: "Yea"<br/>           P1: "So it causes the push and the pull of the magnet, right?"<br/>           P2: "I guess."<br/>           [P1 looking at the AR magnetic field while explaining]<br/>           P1: "The magnetic field is inverted all the time. In order to have music, they change very quickly between forward and backward."</p> | <p>P2: "How do I use this?" [P2 trying to use a compass]<br/>           P1: "This is north, [P1 moving a compass to the other end] this is south. Let me try one second. The moment you come outside of the membrane, it's north. When it is near the membrane, it is south. So north to south, that's how the magnet direction goes."<br/>           [Using compass then iconic hand gesture]<br/>           P2: "How about this part?" [P2 pointing to the coils]<br/>           P1: "This is where they produce electricity current, it starts from the magnetic membrane, this is where the south and north pole comes in."<br/>           P1: "When I push a forward current, it starts at south pole and ends at north pole. And the backward current is north to south."</p> |

## Discussion

In this qualitative analysis, we found that participants using augmented reality exhibited a ‘tunnel vision’ effect, where they spent most of their time utilizing the AR system without leveraging other aids available to them (ex: compass, wall poster), in contrast to the NonAR participants who frequently used external aids. This is a possible explanation for why the AR participants showed higher learning gains in concepts requiring visualization, such as understanding magnetic fields, whereas participants without AR showed higher learning gains in other physical concepts such as relationship between magnetic field and movement. It is possible that AR participants focused strongly on the visual experience and ignored physical aids, while the NonAR focused on physical aids, which in turn increased their awareness of physical effects in the learning experience. Therefore, AR experiences might be a “double-edged sword”: the spatial-temporal contiguity affordance of AR (Radu, 2014) can help participants comprehend complex co-located visual representations, but also impede them from using physical tools and learning about non-visual aspects of the system. In other words, AR representations might be beneficial for learning visual concepts, but detrimental to acquiring kinesthetic knowledge – because users’ attention is so strongly drawn to the “holograms” provided by the AR system, leading them to neglect other sources of information.

Furthermore, our qualitative analysis suggests that AR representations may create a false impression of understanding of the concepts conveyed by the system. In our study, AR participants stopped focusing on the task much faster than NonAR participants; additionally, this effect was stronger for AR groups with low learning gains. This observation is supported by the findings from our larger sample, where we found that AR groups showed higher engagement and beliefs about self-efficacy than compared to NonAR groups (Figure 1), even if learning gains were sometimes worse than NonAR groups (Radu & Schneider, 2019). Furthermore, it may be that NonAR groups persisted with the activity for an extended period due to their need to create representations (such as by drawing, using iconic gestures, or using external aids). This suggests that when participants lack easily accessible information through tools such as AR, the extra effort caused them to engage longer with the content and potentially think more critically. In contrast, AR may give people a false sense of confidence.

Finally, our analysis shows that teaching moments were shorter in the AR groups. When participants had AR visualizations available, communication was more efficient as participants could point or simply refer to an AR visual representation. In contrast, NonAR groups (which lacked the visual representations), had to spend time describing invisible phenomena or generate their own representations by using iconic gestures or drawings. They frequently ran out of time while having to make use of various aids and methods to produce their representation. This aligns with results from Ainsworth (2008) which suggest that to use representational learning tools effectively, learners must be able to first understand the function of the representation as well as how the subject relates to the representation. However, the shorter teaching time with deictic gesture may not necessarily be superior to spending longer time teaching with iconic gesture. Alibali et al (2012) discuss how different gestures serve different purposes for learning and teaching—deictic gesture connects cognition to the physical environment, while iconic gesture represents mental images, and these may correspond to different types of learning.

In future work, we are planning to further investigate why participants in AR spent less time on specific tasks and how different conditions influenced collaborative interactions. To achieve that, we are considering to shrink our video observation time frames from 30 sec. to 15 sec. as we may observe more detailed behavior in shorter intervals. We also plan to revise our coding scheme (communication types) to be able to differentiate degrees in communication (e.g., verbal vs. non-verbal when exploring). Additionally, we are working on augmenting qualitative observations with other process metrics collected from the study (i.e., data from electrodermal wristbands and motion sensors).

## Conclusions

We qualitatively compared video observations of students learning with and without AR visualizations. We found that AR participants learned more about visual concepts but less about non-visual content, stopped exploring the system quicker than NonAR participants, used less aids in exploration and teaching, and spent less time in teaching their collaborators. Those findings suggest that while there might be opportunities to design AR applications in education, there are drawbacks associated with the use of this technology (e.g., the “tunnel vision” effect described above). In summary, the qualitative findings presented in this paper shed new lights on the affordances of AR technology and provides a critical analysis of the use of augmented reality for co-located collaborative learning activities.

## Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 1748093. We would like to thank all our study participants, the anonymous reviewers, the Harvard Decision Science Lab, and all the dedicated research assistants who helped in the research data collection, and our colleague Tonya Bryant who completed the interrater reliability for this study.

## References

- Ainsworth, S. (2008). The educational value of multiple-representations when learning complex scientific concepts. In Gilbert, J. (Ed.), Reiner, M. (Ed.), & Nakleh, M. (Ed.), *Visualization: Theory and practice in science education*. (pp.191-208) Dordrecht: Springer.
- Alibali, M. W., & Nathan, M. J. (2012). Embodiment in Mathematics Teaching and Learning: Evidence From Learners' and Teachers' Gestures. *Journal of the Learning Sciences*, 21(2), 247–286. <https://doi.org/10.1080/10508406.2011.611446>
- Alibali, M. W., Young, A. G., Crooks, N. M., Yeo, A., Wolfgram, M. S., Ledesma, I. M., ... Knuth, E. J. (2013). Students learn more when their teacher has learned to gesture effectively. *Gesture*, 13(2), 210–233. <https://doi.org/10.1075/gest.13.2.05ali>
- Bujak, K. R., Radu, I., Catrambone, R., MacIntyre, B., Zheng, R., & Golubski, G. (2013). A psychological perspective on augmented reality in the mathematics classroom. *Computers and Education*, 68, 536–544. <https://doi.org/10.1016/j.compedu.2013.02.017>
- Cai, S., Chiang, F. K., Sun, Y., Lin, C., & Lee, J. J. (2017). Applications of augmented reality-based natural interactive learning in magnetic field instruction. *Interactive Learning Environments*, 25(6), 778–791. <https://doi.org/10.1080/10494820.2016.1181094>
- Enyedy, N., Danish, J. A., Delacruz, G., & Kumar, M. (2012). Learning physics through play in an augmented reality environment. *International Journal of Computer-Supported Collaborative Learning* (Vol. 7). <https://doi.org/10.1007/s11412-012-9150-3>
- Hill, M., & Sharma, M. D. (2015). Students' representational fluency at university: A cross-sectional measure of how multiple representations are used by physics students Using the representational fluency survey. *Eurasia Journal of Mathematics, Science and Technology Education*, 11(6), 1633–1655. <https://doi.org/10.12973/eurasia.2015.1427a>
- Kohl, P. B., & Finkelstein, N. D. (2007). Expert and novice use of multiple representations during physics problem solving. *AIP Conference Proceedings*, 951, 132–135. <https://doi.org/10.1016/j.apm.2016.02.027>
- Radu, I. (2014). Augmented reality in education: A meta-review and cross-media analysis. *Personal and Ubiquitous Computing*, 18(6), 1533–1543. <https://doi.org/10.1007/s00779-013-0747-y>
- Radu, I. & Schneider, B. (2019). What Can We Learn from Augmented Reality (AR)? Benefits and Drawbacks of AR for Inquiry-based Learning of Physics. In *2019 CHI Conference on Human Factors in Computing Systems Proceedings* (CHI 2019), May 4–9, 2019, Glastow, Scotland, UK. ACM, New York, NY, USA. 12 pages. <https://doi.org/10.1145/3290605.3300774>
- Roth, W.-M. (2001). Gestures: Their Role in Teaching and Learning. *Review of Educational Research*, 71(3), 365-392. <https://doi.org/10.3102/00346543071003365>
- Suyatna, A., Anggraini, D., Agustina, D., & Widyastuti, D. (2017). The role of visual representation in physics learning: Dynamic versus static visualization. *Journal of Physics: Conference Series*, 909(1). <https://doi.org/10.1088/1742-6596/909/1/012048>
- Wu, H. K., Lee, S. W. Y., Chang, H. Y., & Liang, J. C. (2013). Current status, opportunities and challenges of augmented reality in education. *Computers and Education*, 62, 41–49. <https://doi.org/10.1016/j.compedu.2012.10.024>
- Zacharia, Z. C., & Olympiou, G. (2011). Physical versus virtual manipulative experimentation in physics learning. *Learning and Instruction*, 21(3), 317–331. <https://doi.org/10.1016/j.learninstruc.2010.03.001>

# Developing a Model of Collaborative Learning With Minecraft for Social Studies Classrooms Using Role-play Theory and Practice

Anders I. Mørch, University of Oslo, Anders.Morch@iped.uio.no  
Louise Mifsud, Oslo Metropolitan University, Louise.Mifsud@oslomet.no  
Siv Eie, Oslo Metropolitan University, Siv.Eie@oslomet.no

**Abstract:** Minecraft is a multi-user block-building game and 3D virtual world for creating visual structures. We report on two three-week collaborative teaching experiences in a teacher education program in social studies where we experimented with different ways of intertwining block building and role-playing in Minecraft. We created a model for integrating collaborative learning and virtual worlds in social studies classrooms, referred to as collaborative knowledge adaptation. The model consists of three levels: 1) introduction (teacher introducing topic for learning), 2) reconstruction (building in Minecraft and creating a script for role-play), and 3) transformation (acting out role-play and producing a video). Data were collected through interviews, an open-ended questionnaire, and product (video) analysis. We focus our analysis on generic skills interwoven with domain-specific skills and three levels of intersubjectivity toward a shared knowledge object.

## Introduction

Minecraft has been rated the most popular digital game among 9–14-year-old children in Norway (girls/boys 9–11 and in the top five for girls/boys 12–14), but teachers' perception of Minecraft as a game and not a learning tool poses a challenge. This challenge is addressed by a new model for using Minecraft in teacher education, which manages to meet the students in their own arena while fulfilling the requirements of the curriculum. We took advantage of the popularity of the game, opportunities for collaboration and creativity enabled by block building and role-playing, and imposed constraints through knowledge-based activities before, after and in parallel with using Minecraft. Our case is a teacher education program in social studies. Research indicates that three-dimensional (3D) virtual worlds can be integrated into a teacher education program to provide preservice teachers with the experiences needed to apply teaching skills in real school settings (Muir, Allen, Rayner, & Cleland, 2013; Mørch, Hartley & Caruso, 2015).

Minecraft is a sandbox game, which means that users in the game interact by creating visual structures using basic blocks (modeled after 1m<sup>3</sup> physical construction blocks and in analogy to a child playing in a sandbox) and communicate in parallel, mainly by chat but also by voice via other applications. Children are attracted to Minecraft for several reasons; they mine building blocks and craft tools, combat monsters, and collaborate with peers and allies in one of four different gameplay modes. We used the version of the game called Minecraft Education Edition in creative mode, with many of the gaming elements turned off to emphasize the elementary design acts of placing and breaking blocks in order to accomplish more advanced design tasks, such as reconstructing symbolic buildings of a society (e.g., governmental, industrial, cultural, religious, historical). Role-playing was the final stage of our learning activity and occurred inside or outside the buildings.

Engagement and motivation in learning are often listed as the strengths of using games in education (Gros, 2007; Pivec, 2007). Taking advantage of this potential could be the first step toward helping teachers to bridge the gap between students' life worlds and schools' curricular goals. Next, collaboration and problem solving are important behaviors associated with children's digital game playing (Danby et al., 2018), which have been highlighted as important skills for the 21st century that we refer to as generic skills (Mørch, Eie, & Mifsud, 2018). Students combine generic and domain-specific skills when they learn in Minecraft. The aim of our research is to understand how student teachers organize and perceive the value of generic skills in specific subject areas, and how both types of skills can be practiced in the same lesson and in the same digital learning environment. The most prominent generic skills practiced in our study are information seeking, collaboration, and creativity. Other types of generic skills are communication, critical thinking, decision-making, reading, writing, computational skills, adaptability, personal development, and group effectiveness (Binkley et al., 2012).

Minecraft's technological and pedagogical affordances support a broad range of domain-specific skills, including mathematics, chemistry, computer science, history, and social studies (Nebel, Schneider, & Rey, 2016). The students can both visualize and concretize concepts. For example, the constructive play features in Minecraft inherited from wooden blocks, jigsaw puzzles, and Lego bricks can improve children's spatial abilities (Caldera et al., 1999). Furthermore, Minecraft includes tools that enable children to learn programming by introducing it as one of several techniques for solving practical problems in the game environment and

enhancing the game experience. Finally, research has shown that virtual worlds can motivate children to learn specific concepts through immersion and play, which means that they become involved in a subjective experience of concept understanding that leads to a feeling of participating in a comprehensive, realistic experience (Dede, 2009). The domain (subject area) we focus on in this study is social studies.

Minecraft and related virtual worlds and block-building games are suitable for a variety of teaching methods, including scenario-based learning, incremental and iterative activities, and role-playing (Prasolova-Førland et al., 2013; Westera, Nadolski, Hummel, & Wopereis, 2008). In role-play activities, participants are involved in “as-if” or simulated actions to approximate aspects of a real-life situation that is problematic, impractical, impossible, expensive, or risky to carry out in the real world (Yardley-Matwiejczuk, 1997). Students often blur the distinction between themselves and avatars when they immerse themselves in a 3D learning environment, which makes 3D virtual worlds potential sites for embodied and extended cognition (Pasfield-Neofitou, Huang & Grant, 2015). Educators have used role-play in many different application domains, including healthcare, therapy, organizational change, crisis management, military training, and education (Mørch et al., 2015; Prasolova-Førland et al., 2013; Westera et al., 2008). For example, Mørch et al. (2015) studied collaborative learning and role-play in special education in Second Life and found that combining abstract (conceptual) and concrete (hands-on) learning activities in multiple rounds and debriefing sessions were useful for grounding theoretical concepts in concrete experience. We address the combination of abstract and concrete learning activities in our study. We chose Minecraft instead of Second Life, as Minecraft is better adapted to schools.

Role-play and virtual worlds offer interesting opportunities for computer supported collaborative learning (CSCL) research (Jamaludin, Chee, & Ho, 2009). For example, embodied, enactive, extended, and embedded (4E) learning has been proposed as the theme of the CSCL 2019 conference, combining these elements in collaborative settings. Historical re-enactment is an educational or entertainment activity in which people follow a plan to recreate aspects of a historical event or period. In the study presented here this historical enactment is transferred to Minecraft. Our approach builds on ideas of collaborative knowledge construction (Roschelle & Teasley, 1995; Stahl, 2006), particularly knowledge construction achieved by intersubjectivity and meaning-making (Arnseth & Solheim, 2002; Baker et al., 1999; Suthers, 2006), which can be regarded as a dynamically changing context (Suthers, 2006) toward a shared object (Fugelli, Lahn, & Mørch, 2013), in our case a domain-specific knowledge object. Achieving and maintaining intersubjectivity requires a repertoire of conversational acts like “uptake” to reuse and modify prior contributions in conversation (Medina & Suthers, 2013), as well as models for researchers to analyze and design CSCL tasks at different levels of intersubjectivity (Baker et al., 1999) and guidelines for educators to organize their teaching designs (Matusov, 2000).

This study will therefore address the following research questions:

- How are generic and domain-specific skills practiced in the different learning activities?
- How are intersubjectivity and role-play relevant in learning social studies?

The rest of the paper is organized as follows. First, we describe our theoretical framework and the model we partially derived from the framework. Next, we present the methods we have used for data collection and analysis. Then, we present, analyze, and discuss our data and findings and show how the empirical findings informed the model in important ways.

## **Theoretical framework: Role-play and prolepsis-driven intersubjectivity**

The idea of role-play in learning is to foreground envisionment and exploration of alternative courses of action (Yardley-Matwiejczuk, 1997). For example, a child will envision different scenarios for how she will act in an upcoming school presentation by imagining her audience in front of her in the mirror or in her dreams. When this inner and private (non-observable) hypothetical situation is turned into an observable and organized activity, we have a role-play situation, and new opportunities for teaching and learning occur. Vygotsky (1978) suggested that play is a leading factor in human development, stating that “In play a child always behaves beyond his average age, above his daily behavior; in play it is though he were a head taller than himself” (p. 102). The imaginary situations that children engage in when they play are concrete while involving rules and roles, thus providing a means to iterate between concrete and abstract behavior, which we argue is essential for the effective use of role-play in education. Furthermore, Mead (1932) developed a philosophy of experience, with concepts such as emergence, temporality, and consciousness, in which the past and future are viewed through the lens of the present in interactions with others. Mead (1932) emphasized the novel character latent in both the present and the past (e.g., that historical understandings may involve anticipatory cues that may lead to transformations that in turn may lead to new insights).



Role-play conversation is different from naturally occurring dialog in that it is scripted and staged (Yardley-Matwiejczuk, 1997). CSCL researchers have adopted the term script for organizing collaborative learning activities (Cesareni, Cacciamani, & Fujita, 2016). We use script in a slightly different way to mean the written text of a roleplay in film production and connected with terms such as scenes, scripts, props, and prolepsis. We draw on the work of Rommetveit (1976) who developed a prolepsis-driven approach to intersubjectivity. According to Rommetveit, intersubjectivity is a temporarily sustained and partially shared social world that depends on access to historical information (common pre-understanding). Participants in conversation collaboratively construct knowledge by expanding a joint space of intersubjectivity, and a technique for this is to issue anticipatory cues (shared prolepsis), such as subtle utterances, incomplete sentences, or deliberately inserted cues. An incomplete utterance invites the listener to actively participate in the co-construction of an expanded intersubjective space to fill in the missing parts (Rommetveit, 1976; Fugelli, Lahn & Mørch, 2013). An example from our case is providing the role-players with an opportunity to understand a historical event in a contemporary context through an utterance, which points toward a future event outside the context that is shared by the participants, thus making the learning episode more meaningful to them.

The theoretical framework helped us construct a model of collaborative knowledge adaptation for the implementation of virtual worlds in social studies classrooms (Table 1). The columns in the table are connected through three crosscutting themes (introduction, reconstruction, transformation) and are either theoretically motivated (columns 2&3), based on empirical findings (columns 1&5), or informed by both theory and data (column 4).

**Table 1: Collaborative knowledge adaptation model for integrating virtual worlds in classroom practice**

|                | Activity   | Rationale (metaphor of learning)                 | Temporal orientation (Mead, 1932)  | Shared know-ledge object (intersubjectivity)  | Type of skill (primary / secondary)               |
|----------------|--|--|--|---|---|
| Introduction   | Teacher lecturing; students gathering information            | Acquisition (Sfard, 1998)                        | <i>Past:</i> Retrospective views of topic to be learned                          | <i>Vague:</i> Teacher centered; students with different prior experiences   | Domain-specific (primary)                         |
| Reconstruction | Collaborative design in Minecraft and roleplay scriptwriting | Participation (Sfard, 1998)                      | <i>Present:</i> Create new experiences together                                  | <i>Fragmented:</i> Minecraft building vs. role-play script  | Generic (primary) and Domain-specific (secondary) |
| Transformation | Role-playing and video recording                             | Knowledge creation (Paavola & Hakkarainen, 2005) | <i>Emergent:</i> Potential to connect an historical event with an upcoming event | <i>Focused:</i> Enacting an historical event while immersed in a virtual world learning activity (video recorded) | Domain-specific (primary) and Generic (secondary) |

The model is the result of the first iteration of a design experiment. We used part of the model to organize the second iteration; and data we analyze in the next section provide contents for the last two columns (RQ 1&2). The rationale for the model is three metaphors of learning (Sfard, 1998; Paavola & Hakkarainen, 2005), which suggest a sequence of increased complexity of collaborative knowledge construction. We call it collaborative knowledge adaptation, as the first level is domain-specific knowledge introduced by teachers.

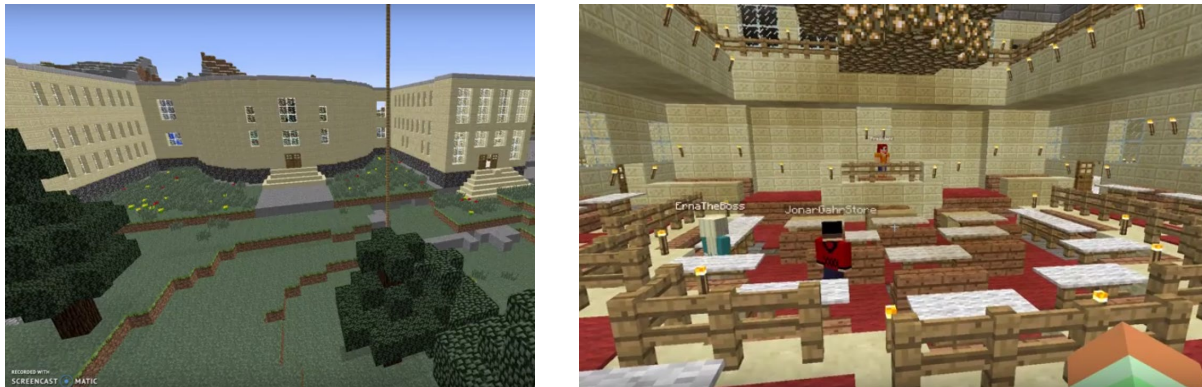
## Methods

The study was organized as an experimental teaching activity between two educational institutions in Norway (A and B) over a three-week period over two semesters (2017 & 2018). Institution A provided educational resources (15 MA students majoring in education acted as tutors) and institution B organized the design experiment as part of a teacher education course (60 BA student teachers majoring in social studies).

Our interventionist approach follows design-based research principles (DBR) (Hoadley, 2002), in which a teacher and two researchers (the authors) collaborate to ‘engineer innovative educational environments and simultaneously conduct experimental studies of those innovations’ (Brown, 1992). DBR as a methodology allows us to focus on contextual aspects that become ‘relevant in the students’ interactions’ (Krange & Ludvigsen, 2009). This means we do not limit ourselves to the local situation of interaction but take the whole

into account in considering how the technical solution provides a new context for interaction. In the first experiment, the activities were building in Minecraft and role-playing with minimal teacher instructions, and in the second experiment the three-level model presented in Table 1 was used to inform our learning design.

The assignment in the first iteration (case 1) was to create a model of the Norwegian parliament and to engage in collaborative learning by role-playing a political decision-making process inside the building based on a script collaboratively created in small groups (Figure 1). The assignment of the second experiment (case 2) was to create a model of one of the historical buildings along the main river of Oslo and create a role-play of a historical event that is said to have happened in and around the building by its owners, tenants, and workers during industrialization, focusing on the working and living conditions of the 19th century (Figure 2).



**Figure 1.** Two video snapshots of Minecraft parliament building (outside and inside, respectively) from case 1.

In both cases, the students were bachelor students, co-located, and worked in groups of three to five individuals. We collected data through an open-ended questionnaire informed by our research questions (N1=37/60; N2=25/40), interviews (in the first case), observation notes, and 25 role-play films (3–5 minutes, both cases). We asked questions about experiences with Minecraft, what activities they found more or less interesting, how they organized collaborative work, the availability of support, would they use Minecraft for own school learning, etc. We thematically coded the answers, following an abductive approach that was partly data-driven and partly theory-driven by our model and research questions (Guest, 2012). We summarize our findings in the next section.



**Figure 2.** Two video snapshots of Minecraft reconstruction of a textile factory building from case 2.

We identified the following themes in our data: 1) time issues, 2) play vs. learning, 3) technical challenges, 4) generic skills interwoven with domain-specific skills, and 5) knowledge sharing (intersubjectivity). We focus on themes 4 and 5 in this paper.

## Results and analysis

A higher percentage of students reported a positive experience using Minecraft in case 2 compared to case 1. Thirteen students (52%) who had not played Minecraft previously indicated that they had satisfactory

knowledge of Minecraft after the three weeks (44% reported previous knowledge of Minecraft). Only one student (4%) claimed that s/he did not master the game after the three-week period (19% in case 1). Findings from interviews in case 1 (Mørch et al., 2018) showed that the student teachers perceive the use of Minecraft in their teaching as a possible threat to domain-specific (social studies) skills practice. However, they consider generic skills important and see the potential in Minecraft as a new way of reaching their students. Our experiences from case 1 made clear to us that the two types of skill (domain specific and generic) were not well integrated in the learning design.

We took the lessons learned from case 1 and started with a domain-specific context on which to base building in Minecraft. The teacher gave a lecture to introduce the social studies topic to the student teachers: the industrial breakthrough in Oslo in the mid-19th century. They were asked to search for information in two ways: through a given list of sources and encouraged to search for additional sources independently. The student teachers were also given an introduction to Minecraft and the generic skills emphasized in the curriculum (information seeking, communication, collaboration). This was followed by an excursion to the river and the industrial area.

## Thematic analysis

To address the two research questions, we give examples from our data showing how generic skills are interwoven with domain-specific skills and how intersubjectivity (knowledge sharing) is achieved during the course of the activity at three levels we have labeled introduction, reconstruction, and transformation.

### Introduction

The students worked in groups after the initial orientation by the teacher. They searched for information through books and the Internet to understand the task. One student elaborated on the process: “We collaborated on finding images of the building from all angles and had a discussion at the physical site” (Student 9). During the excursion to the river and the industrial area, the students measured the perimeter of the building and took photos as preparatory work for the reconstruction to follow. This part of the activity was mainly about acquiring domain-specific (social studies, historical) knowledge and using information seeking as the main generic skill (excluding the basic skills of reading and writing). One student said, “[It took us] around three to four hours to find information about the building, its history, and activity” (Student 18).

Intersubjectivity was not achieved at this level, which we call the “vague object.” Based on our observations, the students, despite knowing of or about each other from previous classes and course assignments, were just beginning to work together on this task. The group efforts included gathering information from various sources (books, Internet texts, pictures of buildings) and preparing the group members for the next level of the activity.

### Reconstruction

The most prominent observed skills practiced at this level were collaboration, cooperation, and design. The students divided the work based on their interests and found that building and creating a role-play scenario were equally important. In their responses to the questionnaire, the students reported that two to three individuals from each group built in Minecraft while two to three wrote the script. As one student noted, “We collaborated well, and split the [task] into different parts. We all contributed toward both building and the role-play [script] but focused on different subtasks” (Student 9). The two subtasks (building and scriptwriting) were interdependent because Minecraft structures provided the scenery for the role-play script and the script defined actions and interactions in the buildings. The students liked this way of working, as it allowed them to be creative and focus on an area they had experience with or interest in while simultaneously contributing to the common goal. As one student put it, they had “the possibility to create something, be creative, and learn to collaborate in groups. The pupils who have problems at school can shine in such an arena” (Student 7). Both tasks made use of knowledge acquired from phase 1, but during role-play planning, the students had to find more detailed information about the historical event so that their re-enactment would be recognizable to viewers outside the group.

Our observation notes and questionnaire data show that two strands of shared knowledge were established at this level: one associated with building in Minecraft and the other scenario development. However, they were not yet integrated and we characterize intersubjectivity as a fragmented knowledge object.

### Transformation

In this phase, the students used what they had created in the previous phase through a roleplay (historical re-enactment) set in the given time epoch (mid-19th century work and life in Norway). The generic skills of

creativity, communication, and collaboration were the most prominent ones observed at this level. For example, creativity had the effect of generating enthusiasm not only within the group but also among the audience, who later watched the video in a plenum session, or as one student put it, “A more enjoyable variation in role-play where pupils can practice other skills such as filmmaking, etc. ... In addition [the pupils] can practice collaboration in film creation” (Student 7). The use of shared prolepsis as a film technique for expanding the space of intersubjectivity among a group of interlocutors and viewers was observed in a few of the videos and is illustrated by the following example:

| Turn | Actor    | Utterance  |
|------|----------|--|
| 1    | Owner    | Yes, indeed, that is a very nice bridge. I will of course maintain it in a good way. Hmm, I think it would be a good idea to introduce a <u>toll</u> collection system, so I can finance maintenance work. |
| 2    | Attorney | <u>Toll</u> ?  |
| 3    | Owner    | Yes, it will be the first toll road in Norway, and with regards to this bridge, which you say is the main crossing between here and Akershus Fortress, I think it's a good idea.                           |



Figure 3. Role-play scene/Minecraft video image.

The extract is part of a dialog between two interlocutors (Owner & Attorney), the owner of the property on which a bridge is located and a legal representative of the government, regarding who should pay for maintenance of the bridge, as the Owner has newly purchased the property. During the exchange, the Owner suggests imposing a toll for crossing the bridge to finance maintenance costs, which comes as a surprise to the Attorney since it is one of a few passage points between west and east in Oslo. The discussion is interesting in that it shows an important event in history, but, by using the word “toll,” the exchange not only takes up and elaborates previous information, but spawns a new line of discourse, connecting the conversation to another event, a current debate in Norway regarding who should finance a new road system around Oslo and other cities (end users vs. central authorities). The students are doing more than recreating history; they are pointing toward a future event they can relate to, which is why we have labeled this level of intersubjectivity *transformation*. It provides a way for students to personalize their collaborative learning efforts in creative ways. Five out of 25 videos included proleptic instances in their role-plays, although there were no instructions to do so. We discuss the implications below.

The two strands of knowledge (Minecraft building and scenario development) are integrated at this level in the three- to five-minute role-play, which are recorded on video and called a focused knowledge object.

## General discussion

We discuss our findings in terms of the research questions we raised in the introduction.

### How are generic and domain-specific skills practiced in the different learning activities?

One of our aims has been to understand how student teachers organize and perceive the value of generic skills in specific subject areas. We addressed this through a model of collaborative knowledge adaptation (Table 1). This model engages the students at different levels of collaborative learning. For example, developing the role-play required in-depth studies of historical sources, which put domain-specific skills practice in the foreground (level 1). Throughout the planning of the role-play, the students had to take a detour and pose “as-if” questions to envision working conditions, class struggles, labor strikes, or rivalries between factory owners. Generic skills (Binkley et al., 2012; Danby et al., 2018) thus became important for the elaboration of domain-specific skills practice: reconstructing historical events through collaborative design in Minecraft and roleplay script writing together (level 2: generic skills practice in foreground and domain-specific in background) and enacting the historical knowledge in creative ways in the role-plays with generic skills in the background (level 3). The data from the open-ended questionnaire indicate that all students found it useful to start with domain-specific skills

and that both the creation of buildings in Minecraft and role-plays (developing and enacting) were equally important.

### How are intersubjectivity and role-play relevant in learning social studies?

Historical knowledge and intersubjectivity have several common features. For example, they are temporary processes that must be periodically updated to sustain, and they are never complete, nor fully shared. Moreover, there are multiple didactical strengths in creating and playing out a historical role-play at different temporal levels. First, the students get a “current” perspective rather than a retrospective. While role-playing, the students get involved as actors and are thus enabled to see history through multiple perspectives and to develop an understanding of historical presentations as constructs (Seixas & Morton, 2013). Second, through role-play the students are creating micro-stories about humans and their environments in the past. These micro-stories can help to develop the students’ historical empathy. The interplay between the micro- and macro-stories will enhance the historical overview and provide a deeper understanding of the past (Kvande & Naastad, 2013).

Matusov (2001) suggested three aspects of intersubjectivity when developing a guide for teaching design: 1) intersubjectivity as having something in common, 2) intersubjectivity as coordination of participants’ contributions, and 3) intersubjectivity as human agency. Our model is inspired in part by these aspects, but it differs in that we make use of role-play and building in a virtual world. Our model can also be compared with the five cycles of enactive role-play sessions proposed by Jamaludin, Chee, and Ho (2009). However, their domain is different from ours, debating moral issues as argumentative dialog rather than techniques for expansion of intersubjectivity. Medina and Suthers (2013) suggested “uptake” as an analytic approach for doing CCSL research. We have expanded the repertoire with “prolepsis,” as a way to project a conversation forward. Our approach builds on the framework of Fugelli, Lahn, and Mørch (2013), who conceptualize intersubjectivity as a knowledge object trajectory from incomplete to more complete. We propose three key events along this trajectory (vague, fragmented, focused), aligned with temporal orientations: past, present, and future (emergent).

### Conclusions, shortcomings, and directions for further work

Our prior experience shows that student teachers tend to value domain-specific skills over generic skills. Therefore, models are needed that emphasize and integrate both types of skills. Towards that end, we have developed the collaborative knowledge adaptation model for integrating virtual worlds in social studies classrooms using a design-based research approach with two iterations, where the results of the first iteration informed the second. The model is based on a combination of empirical findings and theories of role-play and intersubjectivity. The model puts focus on enactment, as in historical re-enactment to address the CSCL 2019 conference theme of 4E learning. We proposed a new conversational act to expand the space of intersubjectivity for small group conversation (shared prolepsis) and gave an example of a new way to use role-play in CSCL.

We plan to continue to develop and test our model in several iterations. We have identified the following shortcomings and directions for further work: (1) The whole activity is time consuming compared to conventional teaching practice; (2) The model is presented as linear, but levels 2 and 3 depend on the previous level, and information may have to be updated (e.g., improving the building based on the role-play); (3) Video recording of collaborative design in Minecraft and script development; (4) Different criteria for evaluating the video products; and (5) The model will undergo four new iterations in higher education and be tested in two experiments in 7th grade classrooms during the 2019-2020 school year, when our student teachers will use Minecraft with 12-year-old pupils in preservice training.

### References

- Arnseth, H.C., & Solheim, I. (2002). Making sense of shared knowledge. Proc. CSCL 2002 (pp. 102-110). Boulder, CO: International Society of the Learning Sciences.
- Baker, M., Hansen, T., Joiner, R., & Traum, D. (1999). The role of grounding in collaborative learning tasks. In P. Dillenbourg (Ed.), *Collaborative learning: Cognitive and computational approaches* (pp. 31-63). New York: Pergamon.
- Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., Rumble, M. (2012). Defining twenty-first century skills. In P. Griffin, B. McGaw & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 17-66). Dordrecht, the Netherlands: Springer.
- Brown, A. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of Learning Sciences*, 2(2), 141-178.
- Caldera, Y.M., Culp, A., O'Brien, M., Truglio, R.T., Alvarez, M., Huston, A.C. (1999). Children's play preferences, construction-play with blocks, and visual-spatial skills: Are they related? *Int. J Behavioral Development* 23(4), 855-872.

- Cesareni, D., Cacciamani, S. & Fujita, N. (2016). Effects of role taking in online writing and reading activities for knowledge building in a blended university course. *Int. J. of Computer-Supported Collaborative Learning*, 11(1), 9-39.
- Danby, S., Evaldsson, A. C., Melander, H., & Aarsand, P. (2018). Situated collaboration and problem solving in young children's digital gameplay. *British Journal of Educational Technology*, 49(5), 959-972.
- Dede, C. (2009). Immersive interfaces for engagement and learning. *Science*, 323(5910), 66-69.
- Fugelli, P., Lahn, L., & Mørch, A.I. (2013). Shared prolepsis and intersubjectivity in open source development: Expansive grounding in distributed work. Proc. CSCW 2013 (pp. 129-144). New York, NY: ACM.
- Gros, B. (2007). Digital games in education. *Journal of Research on Technology in Education*, 40(1), 23-38.
- Guest, G. (2012). *Applied thematic analysis*. Thousand Oaks, California: Sage.
- Hoadley, C. (2002). Creating context: Design-based research in creating and understanding CSCL. Proc. CSCL 2002 (pp. 453-462). Mahwah, NJ: Lawrence Erlbaum Associates.
- Jamaludin, A., Chee, Y.S., & Ho, C.M.L. (2009). Fostering argumentative knowledge construction through enactive role-play in Second Life. *Computers & Education*, 53, 2, 317-329.
- Krange, I., & Ludvigsen, S.R. (2009). The historical and situated nature of design experiments: Implications for data analysis. *Journal of Computer Assisted Learning*, 25(3), pp. 268-279.
- Kvande, L., & Naastad, N. (2013). *Hva skal vi med historie? Historiedidaktikk i teori og praksis*. (Why do we need history? Teaching history in theory and practice). Oslo: Universitetsforlaget. (In Norwegian)
- Matusov, E. (2001). Intersubjectivity as a way of informing teaching design for a community of learners' classroom. *Teaching and Teacher Education* 17, 4, 383-402.
- Mead, G.H. (1932). *The philosophy of the present*. Edited by A.E. Murphy. Chicago and London: The Open Court Publishing Company.
- Medina, R., & Suthers, D. D. (2013). Juxtaposing practice: Uptake as modal transposition. Proc. CSCL 2013 (pp. 328-335). LuLu and Amazon: International Society of the Learning Sciences.
- Mørch, A.I., Eie, S., & Mifsud, L. (2018). Tradeoffs in combining domain-specific and generic skills' practice in Minecraft in social studies in teacher education. Proc. CoPDA 2018 (pp. 44-52), Castiglione della Pescaia, Italy, May 29, 2018 (published at <http://ceur-ws.org/Vol-2101/>).
- Mørch, A.I., Hartley, M.D. & Caruso, V. (2015). Teaching interpersonal problem solving skills using roleplay in a 3D virtual world for special education: A Case study in Second Life. Proc. ICALT 2015 (pp. 464-468). Washington, DC: IEEE Computer Society.
- Pasfield-Neofitou, S., Huang, H., & Grant, S. (2015). Lost in Second Life: Virtual embodiment and language learning via multimodal communication. *Educational Technology Research and Development*, 63(5), 709-726.
- Paavola, S., & Hakkarainen, K. (2005). The knowledge creation metaphor: An emergent epistemological approach to learning. *Science & Education*, 14(6), 535-557.
- Pivec, M. (2007). Play and learn: Potentials of game-based learning. *British Journal of Educational Technology*, 38(3), 387-393.
- Prasolova-Førland, E., Fominykh, M., Darisiro, R., & Mørch, A.I. (2013). Training cultural awareness in military operations in a virtual Afghan village: A methodology for scenario development. Proceedings 46th Hawaii International Conference on System Sciences (pp. 903-912). Washington, DC: IEEE Press.
- Rommetveit, R. (1976). On the architecture of intersubjectivity. In R. Rommetveit & R. M. Blakar (Eds.), *Studies of language, thought and verbal communication* (pp. 147-161). New York: Academic Press.
- Roschelle J., & Teasley S.D. (1995). The construction of shared knowledge in collaborative problem solving. In C. O'Malley (Ed.), *CSCL NATO ASI Series*, vol. 128. Berlin, Heidelberg: Springer.
- Seixas, P. & Morton, T. (2012). *The big six historical thinking concepts*. Toronto, CA: Nelson Publishing.
- Sfard, A. (1998). On two metaphors for learning and the dangers of choosing just one. *Educational Researcher*, 27(2), 4-13.
- Stahl, G. (2006). *Group cognition: Computer support for building collaborative knowledge*. Cambridge, MA: MIT Press.
- Suthers, D.D. (2006). Technology affordances for intersubjective meaning making: A research agenda for CSCL. *Int. J. of Computer-Supported Collaborative Learning*, 1(3), 315-337.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Edited by M. Cole. Cambridge, MA: Harvard University Press.
- Westera, W., Nadolski, R., Hummel, H.G., & Wopereis, I.G. (2008). Serious games for higher education: A framework for reducing design complexity. *J. Comp. Assisted Learning*, 24, 420-432.
- Yardley-Matwiejczuk, K.M. (1997). *Role play: Theory and practice*. London, UK: Sage Publications.



# “Can’t Nobody Floss Like This!”: Exploring Embodied Science Learning in the Third Space

Chris Georgen, Indiana University, cgeorgen@indiana.edu

**Abstract:** I present an out-of-school science study situated in a performing arts center. The design brings together a mixed-reality science learning environment with ensemble dance in a series of embodied learning activities exploring the states of matter. Building on the theoretical framework of the third space, I analyze hybrid embodied movements that highlight the diverse role the body can play in learning science. Additionally, I report on pre- and post-measures indicating participants learned targeted science concepts and hybrid embodiments supported learning one particularly unintuitive concept. Finally, I make a case for the design of embodied learning activities that extend beyond considerations of analogous movements to include local and diverse movements that, if recognized as legitimate, expand possibilities for learning and doing science.

**Keywords:** science, performing arts, technology, elementary education

## Introduction

Educational research has increasingly recognized ways in which the body can play an important role in learning and engaging with disciplined practices (Alibali & Nathan, 2012; Azevedo & Mann, 2018; Goodwin, 2000). More recently, our natural tendency to use the body as a learning resource has been paired with motion-tracking technologies (e.g., mixed-reality environments) in the design of individual and collective instructional contexts that leverage embodiment (Danish et al., 2015; Lindgren & Johnson-Glenberg, 2013). Often these designs focus on some degree of congruency between body movements and the content to be learned, for example pushing arms forward to represent applying force to an object in a Newtonian simulation (Enyedy, Danish, & DeLiema, 2015). However, a focus on cleanly mapping bodily movement to content potentially ignores a diversity and potential hybridity of embodiment and how more informal ways of moving can support learning.

Notions of diversity and hybridity in learning build on the theoretical construct of the third space (Gutiérrez, 2008). The third space highlights the transformative power of hybridity—how informal and formal discourses (e.g., home and school) can become interconnected to productively challenge what counts as knowledge and how knowledge is produced. Hybridity in language practices has been well explored through the lens of the third space (Gutiérrez, Baquedano-López, & Tejada, 1999; Gutiérrez, Rymes, & Larson, 1995). I aim to extend the exploration of hybridity to consider how embodiment can bring together informal and formal ways of moving and knowing (Ma, 2016). In this case, I view hybrid embodiments as those that merge dance-based ways of moving with science. Specifically, I explore how one fun and familiar embodied action (known as “flossing”) can represent a scientific phenomenon to act as a shared resource for participation and sense-making in collective science learning activities. To unpack the role of hybridity and embodiment in learning, I describe the design of a 10-lesson science activity sequence that integrated mixed-reality science learning with ensemble dance. I report on the results of the activity sequence and discuss how hybrid forms of embodiment supported learning by connecting dance movements throughout the intervention to performances on pre- and post-measures. I highlight how the unintuitive scientific concept that solid particles vibrate (i.e., they have high energy but are stationary) is explored that through a variety of ways of moving. I aim to answer the following research questions:

1. How do participants use familiar dance movements to represent and build upon scientific understandings of states of matter?
2. How do these movements support performance on pre- and post-measures?

## Background, design, and community

To illustrate the role of hybridity in embodied learning, I describe the design of a science learning research study that took place in a performing arts center. In contrast to science classrooms, all activity surrounding science learning focuses on local music and dance practices in which the adults and children participate. Hybridity, therefore, is made possible in the ways in which participants make sense of the conceptual practices of science in relation to the local, everyday ways of knowing and doing in dance. The intervention design explicitly connects two thematically and conceptually related ensemble learning activities: mixed-reality science models and dance (Figure 1). Within ensemble settings, individuals learn to take on a role within a system (e.g., a collective dance),

but much of their understanding and performing requires moving and thinking relatively to one another (Ma & Hall, 2018). In science, I explore ensemble learning through what I call collective embodied models. I build on the successful Science Through Technology Enhanced Play (STEP) mixed-reality learning environment that combines motion-tracking technology with a computational simulation (Danish et al., 2015; Danish et al., 2017; Enyedy et al., 2017). STEP allows participants to move across the room pretending to be water particles. The technology tracks their respective movements in real-time to generate a virtual representation of what state of matter those particles would collectively create. Through engaging with and analyzing the simulation, participants learn to coordinate their movement in certain ways to build, test, and refine models that predict and explain the rules behind the particulate nature of matter (Schwarz et al., 2009).

In dance, ensemble learning takes the form of learning and performing choreography. As part of a performing arts program, participants learned, rehearsed, and refined an ensemble dance based on the music and themes of the Disney production *Frozen*. The choreography and materials (e.g., props and costumes) integrate science concepts related to why and how particles move in solid, liquid, and gas. The dance was iteratively rehearsed and refined as participants learned more about the particulate nature of matter and as they practiced an increasingly more complex dance routine. Like the STEP environment, ensemble dance requires working collectively toward representing an idea, feeling, or experience. Since the dance was tied to communicating scientific understanding, it made visible important concepts and supported additional discussion and reflection about what was learned. Yet, the dance was not “scientific.” Though it conveyed aspects of science, the *Frozen* dance was held accountable to dance criteria (e.g., timing, rhythm, presentation, etc.).



Figure 1. Representing ice in STEP (left) and dance (right).

## Methods

### Participants and data sources

Over the course of a summer performing arts program, two groups of participants (n=35, ages 6-8, 34 black or African-American and 1 white) participated in the activity sequence summarized in Table 1. Of these, 18 assented and received parent consent to participate in the study. Participants were divided into two groups. Each group participated in a 45-minute STEP activity, which was immediately followed by *Frozen* dance practice and rehearsal. Each activity had a science learning goal related to rules governing the states of matter, which progressed from macroscopic to microscopic. The STEP activities were facilitated by the research team with support from local staff. The “Frozen Dance” was choreographed a camp instructor. The dance was design to support the ideas from STEP, but maintained the ultimate goal of performing in front of loved ones at the end of camp during a large community event.

Throughout the planning process, the research team met regularly with the dance instructor to discuss the science content, locate overlaps and synergies, and build bridges between the two activity strands. During collaborative planning sessions, the research team and local instructor aimed to connect the science and dance in content, but also thematically through the narrative, characters, and settings from *Frozen*. Connections and similarities between the activities were explicitly communicated to students during activity introductions and debriefs, as well as emergently during key instructional moments. All activities from STEP and dance were video recorded provided consent.



Table 1: Integrated Activity Sequence

| Activity Sequence | STEP Activity   | Dance Activity  | Learning Goal  |
|-------------------|---|---|--|
| Activity 1-2      | Participants embodied energy sources and explored how energy and temperature drive observable changes in state and particle movement. | Participants learned the first phase of their dance focused on representing particle speed and distance during ice to liquid state changes.   | Macro-level: Water exists in different states of matter: solid, liquid, and gas. Water can change states.  |
| Activity 3-4      | Participants took on roles as “energy wands” to explore how energy causes state changes in a snowman, Olaf from <i>Frozen</i> .       | The dance introduced “magic wands” for participants to hold, accenting the dance and representing energy sources in the <i>Frozen</i> dance.  | Macro-level: Energy and temperature produce state changes, i.e., giving energy causes melting and taking away energy causes freezing.  |
| Activity 5-6      | Participants build and test models to explore the relationship between particle distance and speed during state changes.              | The dance introduced “Elsa gloves” for participants to wear, including discussion of how Elsa [the main character in <i>Frozen</i> ] took away energy to make ice in things she touched. Participants learned the second phase of their dance representing liquid to gas state changes. | Micro-level: Develop rules for particle speed and distance in different states (e.g., identify different distances and speeds and associate them with states of matter).   |
| Activity 7-8      | Participants build and test models to explore the relationship between particle attraction during state changes.                      | The dance introduced sticks with streamers on the end for students to hold and represent the chaotic movement of particles in gas.  | Micro-level: Particles in different states of matter are arranged differently. Particles in a solid are patterned, “stuck,” and vibrating. Particles in liquid have no regular arrangement and move closely together. Particles in gas move freely at high speeds. |
| Activity 9-10     | Participants took on roles as water particles inside of containers in the mixed-reality simulation to explore particle behavior.      | Rehearsal for end of camp showcase performance.   | Micro-level to macro-level: Particles in solids always keep their patterned shape, Particles in liquids keep their shape but can move around, and particles in gas move freely and can escape if a container is opened.  |
| Wrap-up           | Post-measures conducted.  | Rehearsal for end of camp showcase performance.   | N/A  |
| Wrap-up           | Post-measures conducted.  | Rehearsal for end of camp showcase performance.   | N/A  |

### Methods: Interaction analysis to investigate hybridity

To understand the role hybrid embodied practices played in learning science content, I used interaction analysis to better understand ways in which participants used dance movements to navigate science concepts and vice versa (Jordan & Henderson, 1995). Due to the focus on the role of hybridity in embodied learning science, I focused video analysis on interactions during STEP activities. Nonetheless, I viewed and content logged the entire dance video corpus to better understand how embodied actions were taken up and used across these activity spaces. Within STEP, I identified constructions of the third space by observing embodiment as it related to the science (e.g., moving like water particles in the simulation) or dance (e.g., moving gracefully like water particles

in dance), focusing on tensions and synergies between these two types of movements. Within these sequences of embodied activity, I transcribed the video verbatim with key attention to formal, informal, and hybrid embodiments. I then categorized and compared these sequences to others to see if any patterns emerged. From there, I flagged moments where the science content was made visible in talk or embodied action to reference with the ways participants performed on pre- and post-measures.

### Methods: Science content knowledge measures

To measure content knowledge, all participants completed a multiple choice pre- and post-test. Additionally, individual pre- and post-interviews were conducted with 18 participants. Interview questions focused on supporting participants in articulating how particles (on the microscopic level) resulted in how matter behaved (on the macroscopic level). I conducted paired t-tests to determine whether students learned about the particulate nature of matter over the course of the activity sequence.

Based on the work of Paik, Kim, Cho, and Park (2004), I coded interviews for reasoning levels about state change mechanisms from circular reasoning (Level 0) or observable macroscopic features (e.g., melting) (Level 1) to unobservable microscopic relationships between kinetic energy and structures of particle movement (Level 4) (Table 2). Coding was not exclusive; any utterance may have more than one code or no codes at all, with the highest level of marking the final code. I transcribed each interview and segments of talk were labeled when participants articulated their reasoning about state changes. For each interview question, codes were applied to each participant utterance, including gestures and embodied actions. I conducted a paired t-test to examine if students' average reasoning level increased from their pre- to post-interview responses. Given my focus on hybridity in movement based on one specific concept, I report on the analysis of one interview question ("How do you think the particles in the ice are behaving?") that specifically addressed participants' experience of hybrid embodiment as it relates to science.

Table 2: State Change Coding Scheme

| <p>Interview Prompt<br/> <i>You might know this already, but things around us, like ice cubes and water, are made up of super tiny pieces called particles. I'm going to ask you some questions about the particles in ice (show student a large ice cube).</i></p> <ol style="list-style-type: none"> <li>How do you think the particles in the ice are behaving?</li> <li>What makes you think so? <i>(If they mention the ice cube slide or moved around, tell them to imagine the ice is completely still.)</i></li> </ol> |  |
|--|--|
| Code   | Example Response   |
| Level 0 - Circular Reasoning: any explanation for why behaviors happen without providing additional information, i.e., things happen because they are just "the way they are."   | "Ice doesn't change shape because it is ice."  |
| Level 1 - Macroscopic explanation (superficial properties): Any explanation that uses observable properties of matter.   | "Ice doesn't change shape because it is hard."   |
| Level 2 - Macroscopic explanation (energy/temperature): Any explanation that uses energy or temperature.   | "Ice doesn't change shape because it is frozen."   |
| Level 3 - Microscopic explanation (single factor): Any explanation that uses the behavior of particles in articulated reasoning, but only gives a single factor in their explanation (e.g., distance, speed, or energy).   | "Ice doesn't change shape because the particles have high attraction."                       |
| Level 4 - Microscopic explanation (complex interactions): Any explanation that uses the behavior of particles in articulated reasoning with multiple connected factors in the explanation (e.g., distance and energy).   | "Ice doesn't change shape because the particles have low energy so the bonds can be strong." |

## Findings

Overall, the integration of STEP and dance afforded participants the opportunity to explore hybrid movements that bridged fun and familiar dance embodiments with scientific ones. This hybridity also appeared to support reasoning across contexts, from dance and science activities to content knowledge post-measures. Throughout the activity sequence, participants moved their bodies in a variety of ways to support reasoning about the science. A key pattern that emerged was the role of the body in reasoning about vibration. To represent ice particles, participants used a ranged of embodied intuitions from more normative scientific embodiments (e.g., standing still and shaking or jumping) to dance-based representations (e.g., doing the “wobble worm” or “slow-mo” dance). In this paper, I focus on “flossing” as an illustrative example of how participants used dance to reason around movement and to accent their scientific understanding in new ways.

## Hybridity in embodied science learning

Because children’s everyday observations of matter tend to lead their assumptions about particle behavior, it can be challenging to support learning about the observable properties of matter as they relate to particulate behavior (Talanquer, 2009). For example, children assume that particles in ice are frozen and therefore do not have energy and cannot move. In fact, solid particles do have energy and vibrate in a rigid lattice structure. While the STEP simulation is useful for supporting understandings about particle distance and movement over space, it is a challenge for young learners to understand that solid particles move but remain stationary. A number of dance-based movements provided a useful resource in exploring ideas related to vibration. The findings of this paper are focused on a specific hybrid embodiment that is taken up by participants that supported robust performances on post-measures related the particles in ice. The following excerpts illustrate one form of hybrid movement that appears to have supported learning about how particles in a solid behave. To illustrate the role of hybrid embodiment in learning about the behavior of particles in ice, Excerpt 1 details a moment when a fun and familiar dance move “the floss” is taken up to represent vibration (Table 3).

Table 3: Excerpt 1

|   |              |  |
|---|--------------|--|
| 1 | Facilitator: | Let’s look on the screen. How are these particles moving?  |
| 2 | Peyton:      | Glitchin’ out.   |
| 3 | Facilitator: | Glitchin’ out. Right. Another way we can say that is vibrating. So, solid particles stay in place and they vibrate ( <i>shaking arms and shoulders</i> ). Everybody stand up and vibrate like a solid particle would. ( <i>Students all stand up and start to shake their bodies</i> ) |
| 4 | PJ:          | Can’t nobody floss like this ( <i>moving arms in front and behind of their body</i> ).   |
| 5 | All:         | ( <i>Other students around PJ begin to laugh and floss</i> )   |
| 6 | Facilitator: | Yeah, PJ! That is a good way to vibrate.   |

During the second week of the activity sequence, students took on roles as energy sources and observe how energy effects particles in different states. Excerpt 1 took place while participants were making observations about how solid particles are moving in the STEP simulation. One participant, Peyton described the particles as “glitchin’ out” (Line 2). This bid was taken up by the facilitator as an opportunity to introduce the concept of vibration and a more normative embodied representation of standing still and shaking arms and shoulders. This embodied movement is transformed by another participant, PJ, who responded by performing the well know dance “the floss” (Line 4). Flossing is a dance move performed by standing in place still while swinging arms, with hands in fists, from the back of the body to the front and moving hips in the opposite direction (Figure 2).

Flossing was made popular on social media and exploded due to its performance in music videos and video games. While flossing was not an explicit aspect of the final choreography in the *Frozen* dance, it was an “insider” movement in the context of this performing arts center, and children could be seen flossing throughout the day (e.g., during free time and lunch). Flossing also often occurred informally during STEP activities that promoted the exploratory use of the body. However, in this instance PJ spontaneously introduced flossing to move like a solid particle in science (i.e., standing in place and vibrating). While not scientifically normative, the move to represent particles with locally recognized dance blended the scientific notion of vibration with a well-known

dance. Moreover, it anchored students' embodiment in a way that legitimately maps to the verbal descriptions of vibrating particles as staying stationary but moving.



Figure 2. Choreographic movements of “the floss.”

After this moment, flossing was taken up as a legitimate way to represent the trajectory of particles in ice during later activities. Beyond the overlaps in embodiment, the “unofficial” movement of “the floss” became central to how students understand particulate behavior. Hybridity both leveraged the prior knowledge and experiences of the participants as well as expanded the possible ways of knowing and moving related to science. For example, Excerpt 2 occurs toward the end the activity sequence. This activity focused on how particles in different states behave given the macroscopic conditions of an environment (e.g., a container with a lid) (Table 4). After engaging with the mixed-reality simulation as particles, the students re-represented their understanding from STEP with an embodied model. During the activity, the participants pretended the classroom was a container and the door was the lid. Beginning as ice particles, the facilitator gave the particles energy until they ran around the room as gas. Once the door was opened, all the participants ran out of the room like gas particles escaping a container.

Table 4: Excerpt 2

|   |              |   |
|---|--------------|---|
| 1 | Facilitator: | If we were particles, what state of matter would we be in right now?    |
| 2 | All:         | Solid!  |
| 3 | Facilitator: | Solid, right. Do solid particles stay totally still?                    |
| 4 | All:         | No.   |
| 5 | Lauren:      | No, they vibrate! <i>((Participants start to shake arms and floss))</i> |

Excerpt 2 begins with a typical instructional move of opening instruction with a content question (Line 1). At once, the participants respond with the correct answer that particles in this formation would create a solid (Line 2; Figure 3). Further probing the participants, the instructor calls for more detailed information about particles, which is returned with another collective response. Immediately after, another participant, Lauren, calls out the answer, which is next embodied by Lauren and her peers with flossing. This moment demonstrates that the original embodied connection persisted over the course of multiple activities, as many of the participants took up flossing to represent vibration in solid particles.



Figure 3. Embodied re-representation of solid particles.

## Learning gains

Paired t-test results of the 10-question multiple choice measure indicate that individual participants significantly improved their general understanding about particle behavior ( $t(17)=-7.071$ ,  $p < 0.001$ ) from pre-test (mean=3.11, SD=1.9) to post-test (mean=6.44, SD=0.85).

Given the specific focus on how hybrid embodied movements supported learning, the analysis of pre- and post-interviews focused on the first question of the interview protocol: How do you think the particles in ice are moving? Paired t-test results of the highest level code applied to student responses to this question indicate that individual participants significantly advanced their reasoning ( $t(17)=-8.416$ ,  $p < 0.001$ ) from pre-test responses (predominately circular reasoning or superficial observations) (mean=0.944, SD=0.53) to post-test responses (single and multifactor micro-level particle behaviors) (mean=2.61, SD=0.77). Generally, in pre-interviews participants focused on macro-level properties of ice to reason around how particles in ice behaved. For example, most students gave the expected response that particles in a solid were frozen together and could not move. In post-interviews, most participants (14/18) said that solid particles vibrate (see Table 5). This result is in marked contrast to previous classroom implementations of the STEP activity sequence focused on modeling and embodied play where only half the students in post-interviews could accurately describe or demonstrate how particles in ice behave (Danish et al., 2015). While these previous implementations supported robust learning gains, participants' beliefs about ice particles persisted despite experiences in the mixed-reality science simulation. Given that pre-interviews in this and similar STEP studies showed that the groups were comparable on their prior knowledge of science, it seems that the dance context contributed to participants' ability to engage with the conceptual knowledge and practices of science. That is, the dance context that supported hybrid embodied movements appears to have supported students in moving beyond descriptions of solid particles based on a surface-level characteristic.

Table 5. Interview Responses

| How do you think the particles in the ice are behaving? | Particles are frozen or don't move | Particles are vibrating | Circular reasoning |
|---|------------------------------------|-------------------------|--------------------|
| Pre-interview   | 13                                 | 0                       | 5                  |
| Post-interview  | 4                                  | 14                      | 0                  |

## Summary and significance

The integrated sequence of STEP and dance activities resulted in both learning gains and opportunities for participants to explore diverse forms of embodiment. By focusing on moments of hybridity, analysis traced specific movements from their spontaneous enactment to their use as a shared resource for sense-making in collective activities and individual assessments. In particular, flossing appears to have supported understanding an unintuitive concept that historically challenged young learners (i.e., that solid particles do in fact move). These outcomes suggest that scientific embodiments can incorporate local ways of knowing and moving and become a powerful resource in meaning-making. These results have implications for researchers and practitioners. First, work should further recognize embodiment as a diverse and hybrid practice. Second, scholars should continue to encourage the co-construction of hybrid learning spaces in a variety of modalities. Third, this study helps to articulate how future designs for embodied activity can benefit from moving beyond considerations of analogous movement (although that is still important) to include a focus on a diversity of movements that, if properly supported, can blend personally-meaningful movements with scientifically normative ones.

Overall, the goal of this paper is to highlight the need to understand the diversity of potential body movements in designing and implementing classroom-based and out-of-school science learning contexts. I argue that these results further support the central role the body plays in learning and exploring concepts. At the same time, it expands notions of embodied cognition from the ways a deliberately cued and technologically-enhanced body might directly interface with science content to include spontaneous new forms of movement that support learning (Lindgren, 2014). Moreover, this work offers new perspectives on hybridity and the third space—that is, how an *embodied* third space can produce and invite new forms of body-based participation for both individual participants and collectives. In sum, these outcomes can continue to advance computer-supported collaborative learning contexts, and in doing so expand possibilities for young learners' future scientific experiences and identification (Bell, Tzou, Bricker, & Baines, 2012).

## References

- Alibali, M. W., & Nathan, M. J. (2012). Embodiment in mathematics teaching and learning: Evidence from learners' and teachers' gestures. *Journal of the learning sciences, 21*(2), 247-286.
- Azevedo, F. S., & Mann, M. J. (2018). Seeing in the dark: Embodied cognition in amateur astronomy practice. *Journal of the Learning Sciences, 27*(1), 89-136.
- Bell, P., Tzou, C., Bricker, L., & Baines, A. D. (2012). Learning in diversities of structures of social practice: Accounting for how, why and where people learn science. *Human Development, 55*(5-6), 269-284.
- Danish, J. A., Enyedy, N., Saleh, A., Lee, C., & Andrade, A. (2015). Science Through Technology Enhanced Play: Designing to Support Reflection Through Play and Embodiment In Lindwall, O., Häkkinen, P., Koschman, T. Tchounikine, P. Ludvigsen, S. (Eds.) (2015). Exploring the Material Conditions of Learning: The Computer Supported Collaborative Learning (CSCL) Conference 2015, Volume 1. Gothenburg, Sweden: The International Society of the Learning Sciences.
- Danish, J., Humburg, M., Saleh, A., Lee, C., Dahn, M., Kiefert, D., & Enyedy, N. (2017). A Socio-Cultural Framework for Embodied Cognition. Paper presented at the Jean Piaget Society, San Francisco, CA.
- Enyedy, N., Danish, J., DeLiema, D., Saleh, A., Lee, C., Morris, N., & Illum, R. (2017). Social Affordances of Mixed Reality Learning Environments: A case from the Science through Technology Enhanced Play project (STEP).
- Goodwin, C. (2000). Action and embodiment within situated human interaction. *Journal of pragmatics, 32*(10), 1489-1522.
- Gutiérrez, K. D. (2008). Developing a sociocritical literacy in the third space. *Reading research quarterly, 43*(2), 148-164.
- Gutiérrez, K. D., Baquedano-López, P., & Tejeda, C. (1999). Rethinking diversity: Hybridity and hybrid language practices in the third space. *Mind, culture, and activity, 6*(4), 286-303.
- Gutierrez, K., Rymes, B., & Larson, J. (1995). Script, counterscript, and underlife in the classroom: James Brown versus Brown v. Board of Education. *Harvard educational review, 65*(3), 445-472.
- Lindgren, R., & Johnson-Glenberg, M. (2013). Emboldened by embodiment: Six precepts for research on embodied learning and mixed reality. *Educational Researcher, 42*(8), 445-452.
- Lindgren, R. (2014). Getting into the cue: Embracing technology-facilitated body movements as a starting point for learning. In *Learning Technologies and the Body* (pp. 51-66). Routledge.
- Jordan, B., & Henderson, A. (1995). Interaction analysis: Foundations and practice. *The journal of the learning sciences, 4*(1), 39-103.
- Ma, J. Y. (2016). Designing disruptions for productive hybridity: The case of walking scale geometry. *Journal of the Learning Sciences, 25*(3), 335-371.
- Ma, J. Y., & Hall, R. (2018). Learning a part together: ensemble learning and infrastructure in a competitive high school marching band. *Instructional Science, 1*-26.
- Paik, S. H., Kim, H. N., Cho, B. K., & Park, J. W. (2004). K-8th grade Korean students' conceptions of 'changes of state' and 'conditions for changes of state'. *International Journal of Science Education, 26*(2), 207-224
- Schwarz, C. V., Reiser, B. J., Davis, E. A., Kenyon, L., Achér, A., Fortus, D., ... & Krajcik, J. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of research in science teaching, 46*(6), 632-654.
- Talanquer, V. (2009). On cognitive constraints and learning progressions: The case of "structure of matter". *International Journal of Science Education, 31*(15), 2123-2136.

# Fostering Collective Knowledge Advancement Through Idea-Friend Maps in a Large Knowledge Building Community

Xueqi Feng, The University of Hong Kong, fengxueqi@hotmail.com  
Jan van Aalst, The University of Hong Kong, vanaalst@hku.hk  
Carol K.K. Chan, The University of Hong Kong, ckkchan@hku.hk  
Yuqin Yang, Central China Normal University, yuqinyang0904@gmail.com

**Abstract:** This study designed and examined a knowledge-building environment that combined a learning analytics tool for creating “Idea-Friend Maps” with Knowledge Forum for Primary Five students. Students in the experimental class (n=53) used the Idea-Friend Map tool, whereas those in the comparison class (n=54) conducted the same knowledge-building activities without the use of analytics tools. Results showed that students in the experimental class showed better conceptual understanding, higher levels of social participation, and more advanced collective knowledge. Analyses were conducted on how students used analytics to support their knowledge building inquiry. These findings suggested that learning analytics can be used to promote knowledge building in large communities.

## Introduction

China is committed to education reform and is open to Western theories that have been developed in the learning sciences (Ryan, 2013). Knowledge building is a major pedagogical approach in the learning sciences community that focuses on theory building and knowledge creation (Scardamalia, 2002). The online platform Knowledge Forum® (KF) was designed to promote collective knowledge advancement by creating an online environment where students can pose problems, propose explanations, test ideas, and conduct sustained pursuit of inquiry to facilitate collective knowledge advancement. Generally, most studies on knowledge-building practice involve smaller-sized classes (15-30 students) (Chen, Scardamalia, & Bereiter, 2015; Oshima et al., 2004; van Aalst & Truong, 2011; Zhang, Scardamalia, Reeve, & Messina, 2009); however, classes in China generally contain 50-60 students, and very few studies involve classes with more than 50 students. Thus, little is known about how to implement knowledge building in such settings, nor about issues of scale in online discourse. How to facilitate large knowledge-building communities engaging in productive discourse is a challenging issue.

Generally, when classes larger than about 20 students, teachers have adopted group-based instruction (van Aalst, 2009) in knowledge building environments. Zhang et al. (2009) examined different social configurations—i.e., fixed groups (students work in fixed groups), interactive groups (interacting groups with cross-group activities), and opportunistic groups (new groups forming and disbanding based on emergent goals)—and found opportunistic groups yielded the best learning outcomes. However, whether opportunistic interaction adapts to large class environments still needs to be verified. Furthermore, when such a community promotes flexible participation and emergent inquiries on KF, it is difficult for students to understand the changing status of collective knowledge at the community level (Kimmerle, Cress, & Hesse, 2007). Therefore, the most urgent challenge is how to design an environment to help large class students identify, from the large volume of online discussion notes, the cutting edge of community knowledge, conduct sustainable problem solving, and structure what the community should investigate. However, the large number of online notes generated by large classes may provide opportunities for educational data mining and learning analytics, as it involves large datasets (Baker & Inventado, 2014).

There is increasing interest in learning analytics that emphasizes informing and empowering learners and instructors to improve their learning processes (Nistor & Hernández-García, 2018). However, much attention in this field involves evaluating what learners have done or predicting what they will do in the future (Siemens & Long, 2011); the area of how analytics are used as part of teaching and learning processes is relatively unexplored (Wise, Vytasek, Hausknecht, & Zhao, 2016). Knowledge Building Discourse Explorer (KBDex) is a novel learning analytic tool designed to investigate student-, keyword-, and discourse-based social networks on KF (Oshima, Oshima, & Matsuzawa, 2012). However, many current studies use KBDex merely as a tool for evaluating learning processes (Ma, Matsuzawa, Chen, & Scardamalia, 2016; Matsuzawa, Oshima, Oshima, & Sakai, 2012; Resendes, Scardamalia, Bereiter, Chen, & Halewood, 2015). Our program aims to design representations of the learning analytics tool KBDex, integrated with relevant social configurations for Grade Five students in large knowledge-building communities, and to examine the designed environment’s impact on students’ learning outcomes and collective knowledge advancement. Our research questions are: (1) Did students in the experimental class have better learning outcomes, greater KF participation, and deeper KB discourse? (2)



How did students in the experimental class engage in collective knowledge advancement, with the support of learning analytics?

## Methods

### Participants

This study was carried out in two Grade Five classes in a primary school in China, with 53 students in the experimental class and 54 in the comparison class. The first author was also the science teacher of the class. These students have had one-year experience in learning using knowledge building with this teacher before this study.

### Pedagogical design

The two classes investigated the topic of *Electricity* on KF (Figure 1) for 9 weeks. In Phase 1 (Weeks 1-3), students in both classes first created KF notes for inquiry, categorized their online notes into seven subtopics, formed learning groups, and chose subtopics according to interests. In Phase 2 (Weeks 4-6), students in both classes interacted with each other in different subtopics, did experiments, and conducted theory-building; however, students in the experimental class used the group-level Idea-friend Maps (IFM) (Figure 2, (a) and (b)), while those in the comparison group did not. In Phase 3 (Weeks 7-9), students in both classes worked in opportunistic groups (students formed/dissolved new groups based on emergent goals/common interests) to contribute to community views (included all the subtopics); however, students in the experimental class used the community-level IFM (Figure 2, (c) and (d)), while those in the comparison group did not.

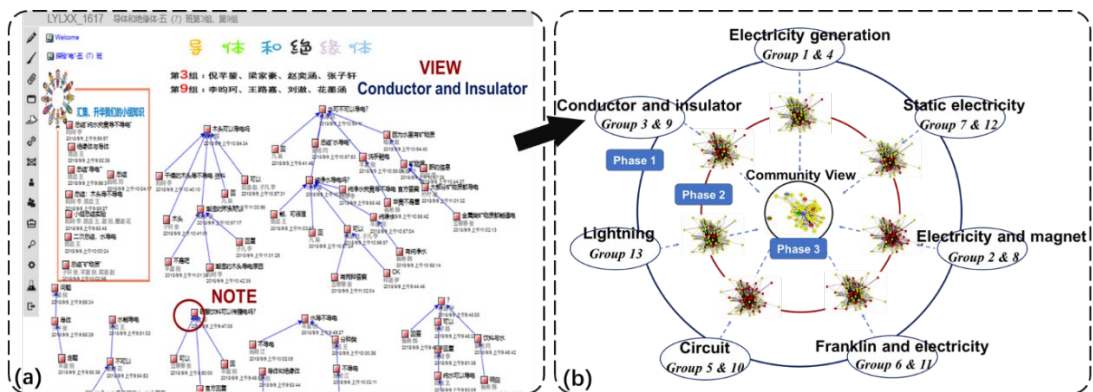


Figure 1. The KF view of *Conductor and Insulator* (a); the experimental pedagogical design (b).

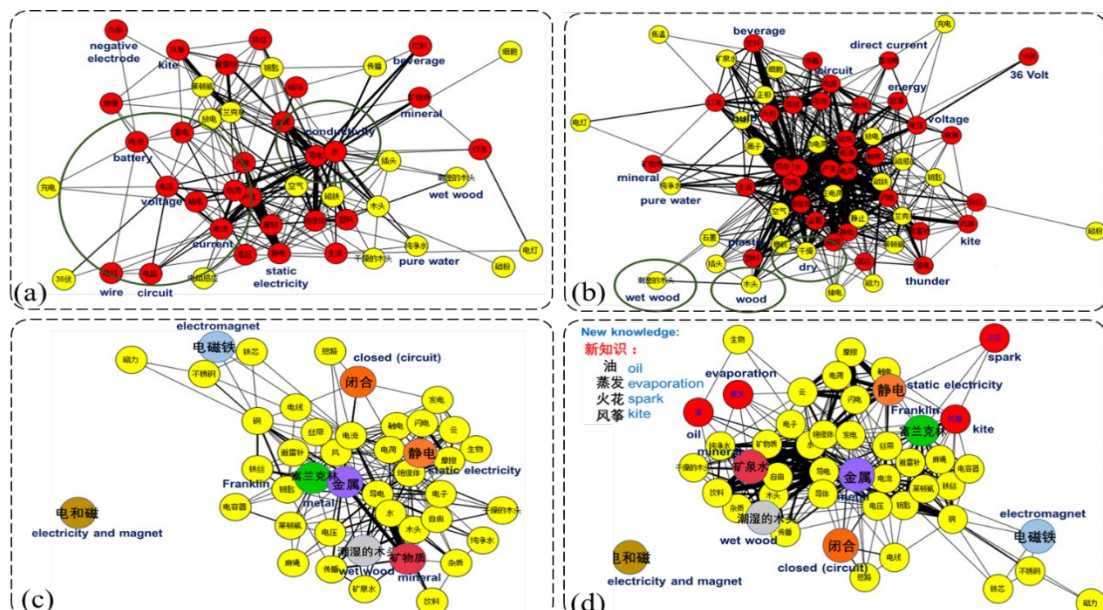


Figure 2. The intervention of IFM: (a) Group 9's group-level IFM in Week 4; (b) Group 5's group-level IFM in Week 5; (c) community-level IFM in Week 8; (d) community-level IFM in Week 9.



## The learning analytic tool (Idea-Friend Map)

The IFM intervention was the word network exported from KBDEX. We designed the group-level IFM to help interactive activities in Phase 2, and the community-level IFM to promote opportunistic knowledge advancement in Phase 3. Figure 2 (a) and (b) show group-level IFM of different groups in different weeks; the red and yellow balls indicate words that had or had not already been discussed by the group, respectively; yellow balls near red balls indicate idea friends. For instance, students in Group 9 had already discussed those red balls words as “wire”, “circuit” and “current” until Week 4 (see Figure 2 (a)), while those yellow balls might be discussed by other groups. Words “dry” and “wood” from Figure 2(b) were “plastic”’s idea friends because they were very close to “plastic”. Students in different groups first identified idea friends, then went to other subtopics that included those ideas friends, enabling them to identify connections among different science concepts, make learning plans, interact with different groups, which would result in the advancement of collective knowledge. Correspondingly, Figure 2 (c) and (d) show community-level IFM during the final weeks of the intervention; balls with different colors (except yellow) indicate key topics identified in the community view, allowing students in opportunistic groups to synthesize different topics, summarize a topic, and create new knowledge with the support of community-level IFM. For example, the green ball with “Franklin” from Figure 2 (c) represented the problem of “Why did Benjamin Franklin not die from his lightning strike?” (see Table 2).

## Data analysis and findings

RQ1: Did students in the experimental class have better learning outcomes, greater KF participation, and deeper KB discourse?

### Changes in science learning outcomes across classes

We analyzed students’ science domain tests to examine the effects of IFM on students’ learning outcomes. There was no significant difference in the pre-test scores of the two classes ( $F(1, 101) = 0.81, p = .372$ ). A two-way (Intervention  $\times$  Time) ANOVA showed a significant main effect of Time ( $F(1, 199) = 145.53, p < .001$ , Partial  $\eta^2 = .42$ ). A significant Time  $\times$  Intervention effect was obtained ( $F(1, 199) = 5.84, p = .017$ , Partial  $\eta^2 = .03$ ), indicating that students in the experimental class improved more on science learning outcomes, compared to students in the comparison class.

### Class differences in KF contribution and interaction

We assessed quantitative data about the number of students’ notes written, notes linked, notes read, scaffolds used, reference used, and views worked on KF, using the analytic toolkit (ATK) (Burtis, 1998). One-way ANOVA was performed to assess whether students’ contributions to the online discourse were different among the classes. There were significantly higher values of notes read ( $F(1, 105) = 10.10, p < .01$ ), views worked ( $F(1, 105) = 10.54, p < .01$ ), and references used ( $F(1, 105) = 13.11, p < .001$ ) in the experimental class. In contrast, there was no significant difference in the notes written, notes linked, and scaffolds used between the two classes ( $p > .1$ ). These results suggest that the IFM intervention helped students to engage in different views, read more notes, and reference useful ideas.

### Class differences in student online knowledge building discourse

We analyzed 588 KF notes from the experimental class and 528 from the comparison class to characterize students’ contributions to knowledge-building discourse. We adopted a framework to code the notes in each interactive view and the community view. The coding schemes included three main categories—*questioning*, *theorizing* and *community*—and corresponding subcategories, and drew upon theoretical frameworks for progressive inquiry (Hakkarainen, 2003) and knowledge creation (Chuy, Zhang, Resendes, Scardamalia, & Bereiter, 2011; Fu, van Aalst, & Chan, 2016). Two raters independently coded about 50% of all the notes. The inter-rater reliability was .94 for *questioning*, .94 for *theorizing*, and .93 for *community* (Cohen’s  $\kappa$ ).

As Table 1 shows, the experimental class created more *theorizing* and *community* notes, while the comparison group tended to raise questions. One-way ANOVA was adopted to further examine the differences between the two classes and found significantly higher values of *improving an explanation* ( $F(1, 14) = 5.88, p < .05$ ), *bridging knowledge* ( $F(1, 14) = 8.30, p < .05$ ), and *synthesis* ( $F(1, 14) = 5.15, p < .05$ ) in the experimental class. We also observed significant higher values for *simple claim* ( $F(1, 14) = 5.88, p < .05$ ) created in the comparison class. The differences in other codes were not significant between the two classes ( $p > .05$ ). These results indicate that, when confronted with new problems, students in the experimental class tended to bridge knowledge through the website, engage in deeper theory-building processes, and synthesize useful ideas, while the comparison group tended to ask more fact-oriented questions and respond with simple claims, rather than seek relative notes or authoritative materials.

Table 1: Number of different categories of questioning, theorizing, and community in interactive views (experimental class/ comparison class)

| View  | Questioning  |                      |                   | Theorizing   |                          |                           | Community                |                   |                 |                    |           |
|-------|--------------|----------------------|-------------------|--------------|--------------------------|---------------------------|--------------------------|-------------------|-----------------|--------------------|-----------|
|       | Fact seeking | Explanati on seeking | Sustained inquiry | Simple claim | Proposing an explanation | Supporting an explanation | Improving an explanation | Shared regulation | Lending support | Bridging knowledge | Synthesis |
| #1    | 7/7          | 15/10                | 8/0               | 7/14         | 29/40                    | 16/9                      | 10/1                     | 8/0               | 0/1             | 15/1               | 2/0       |
| #2    | 8/9          | 9/12                 | 6/3               | 19/24        | 23/16                    | 15/4                      | 8/1                      | 2/2               | 1/1             | 18/9               | 2/1       |
| #3    | 0/5          | 3/7                  | 1/2               | 0/16         | 6/11                     | 4/1                       | 2/1                      | 2/1               | 0/3             | 10/10              | 0/0       |
| #4    | 3/16         | 8/9                  | 6/6               | 12/15        | 15/26                    | 7/9                       | 10/0                     | 4/1               | 2/4             | 6/13               | 1/0       |
| #5    | 8/13         | 7/7                  | 3/3               | 16/34        | 7/7                      | 3/2                       | 1/0                      | 2/2               | 1/1             | 12/5               | 2/1       |
| #6    | 15/19        | 11/12                | 6/7               | 13/26        | 39/16                    | 11/4                      | 10/3                     | 5/3               | 1/2             | 18/7               | 9/1       |
| #7    | 2/10         | 3/14                 | 4/1               | 10/18        | 8/16                     | 0/4                       | 1/0                      | 0/2               | 4/3             | 12/10              | 2/1       |
| #8    | 2/7          | 9/11                 | 19/13             | 1/6          | 38/26                    | 22/10                     | 30/1                     | 1/0               | 0/0             | 15/6               | 9/0       |
| Total | 45/86        | 65/82                | 53/35             | 78/153       | 165/158                  | 78/43                     | 72/7                     | 24/11             | 9/15            | 106/61             | 27/4      |
| Mean  | 5.63/10.75   | 8.13/10.25           | 6.63/4.38         | 9.75/19.13   | 20.63/19.75              | 9.75/5.38                 | 9.00/0.88                | 3.00/1.38         | 1.13/1.88       | 13.25/7.63         | 3.38/0.5  |
| SD    | 4.87/4.86    | 3.98/2.49            | 5.45/4.21         | 6.76/8.61    | 13.66/10.48              | 7.55/3.46                 | 9.43/0.99                | 2.56/1.06         | 1.36/1.36       | 4.10/3.70          | 3.54/0.53 |

Notes: Views #1 to #7 mean subtopics that interactive groups worked on, while view #8 means the community view. #1=Electricity generation; #2=Franklin and electricity; #3=Lightning; #4=Static electricity; #5=Circuit; #6=Conductor and insulator; #7=Electricity and magnet

Table 2: Categories of problems in the community views for both classes

|                    | Static electricity  | Circuit   | Conductivity  | Franklin and electricity   | Electricity and magnet  |
|--------------------|---|---|---|--|---|
| Experimental class | #1 How does lightning form?<br>#2 What is the object can create static electricity? #3 Why is static electricity formed from friction? #4 Can I see current when static electricity is transmitted? | #1 Why does circuit need a closed loop? #2 How does the body make electricity? #3 Why does people might die from an electric shock? | #1 Why is mineral water/beverage/wet wood/metal conductive?                   | #1 Why did Benjamin Franklin not die from his lightning strike? #2 What is the relationship between a lightning rod and insulator? | #1 Why is iron used for the core of an electromagnet? #2 What is the relationship between electricity and magnet? |
| Control class      | #1 How does lightning form?<br>#2 Why is static electricity formed from friction? #3 How to remove static electricity?  | #1 Why does circuit need a closed loop? #2 How is electricity produced? #3 Will a short circuit cause an explosion?                 | #1 Why can't the insulator be conductive? #2 Why is metal/mineral conductive? | #1 Why did Benjamin Franklin not die from his lightning strike?  | #1 What is the relationship between electricity and magnet?   |

## RQ2: How did students in the experimental class engage in collective knowledge advancement with the support of the learning analytics?

### Class differences in interactive views

We compared key interactive views reflecting how student advanced collective knowledge with the support of the group-level IFM. We first quantitative compared the Total Degree of Centrality (TDC) of the KBDEX word network to assess the collective knowledge advancement (Oshima et al., 2012) of the interactive views *Conductor and Insulator* (Figure 3(a)) and *Circuit* (Figure 3(b)). We focused on these two views because they are the most fundamental parts when students discussed *Electricity*. A higher degree centrality means a denser social network. Figure 3 shows the word network's increasing TDC over time, indicating how students collectively worked on key science ideas. In Phase 1, there were no significant differences in the collective knowledge advancement of the two classes in either view ( $p > .5$ ), when they shared the same instruction. However, there were significantly higher TDC values in *Conductor and Insulator* ( $F(1, 4) = 101.86, p = .001$ ) and *Circuit* ( $F(1, 4) = 112.16, p < .001$ ) for the experimental class compared to the comparison class when students used IFM, in Phase 2. We also qualitatively analyzed students' online notes, prompt sheets and experiment reports when they engaged in the interactive views, *Conductor and Insulator* and *Circuit*.

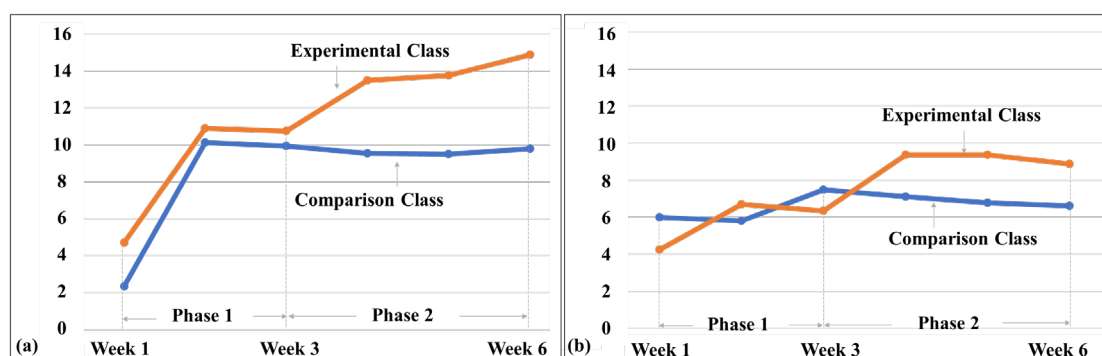


Figure 3. TDC of word network for the *Conductor and Insulator* view (a) and the *Circuit* view (b).

### View: *Conductor and insulator*

We focused on the students from Group 9 in the experimental class who took responsibility for the *Conductor and Insulator* view to clarify how they engaged in collective knowledge advancement with the support of group-level IFM. After posting questions, proposing explanations, and searching authoritative information in the previous three weeks of Phase 1, Group 9 students had a general understanding of conductors and insulators, then they looked at their group IFM in Week 4, they found the keywords related to their topics, such as “conductivity”, occupied the right half of the entire IFM (Figure 2 (a)), while some keywords, such as “battery” and “voltage”, belonged to the *Circuit* group. Therefore, they speculated that, if they wanted to further explore their own topic, they must first understand the basic knowledge of “circuit”. Then they went to the *Circuit* view and managed to design an experiment to verify the elements of a circuit by lighting up a light bulb (Figure 4 (a)) by themselves. Later, they successfully revised their circuit experiment to explore whether water/wood was conductive (Figure 4 (b)). In addition, the students could also spontaneously review their thinking processes and try to explain their experimental phenomena in light of another's perspective. Below is a summary note tracing how they explored the relationship between *Circuit* and their own topic, *Conductor and Insulator*.

**[My idea]** What is the relationship between circuit and conduction? In fact, the most basic thing about conduction is the circuit... To make an object conductive, we must first do a circuit experiment.

☑ “Group summary of experiment” by Group 9) We conducted two experiments: the first was to make the bulb light up. This experiment was a pilot for conductor and insulator... The second experiment, of water's conductivity, was the experiment we originally wanted to explore... Because pure water has no impurities, most of it is not electrically conductive. In mineral water there are impurities, so the light bulb is illuminated when the wire is placed in it, so the two are electrically conductive.

☑ “Reply” by student c726) c726 also said that water can conduct electricity because there are cells inside it. You can understand in this way.

But the *Conductor and Insulator* group in the comparison class tended to ask fact-oriented questions, simply look up information, and copy/paste difficult conceptions from websites. The following is a cluster from the comparison group discussing the same topic (water's conductivity).

[I want to know] What object can be conductive?

[My idea] Water!

[I want to know] Why is water conductive?

[My idea] If all the impurities in the water are removed, the water resistivity will increase up to 18.2 megaohms x cm...

[My idea] Please provide information that we can understand.

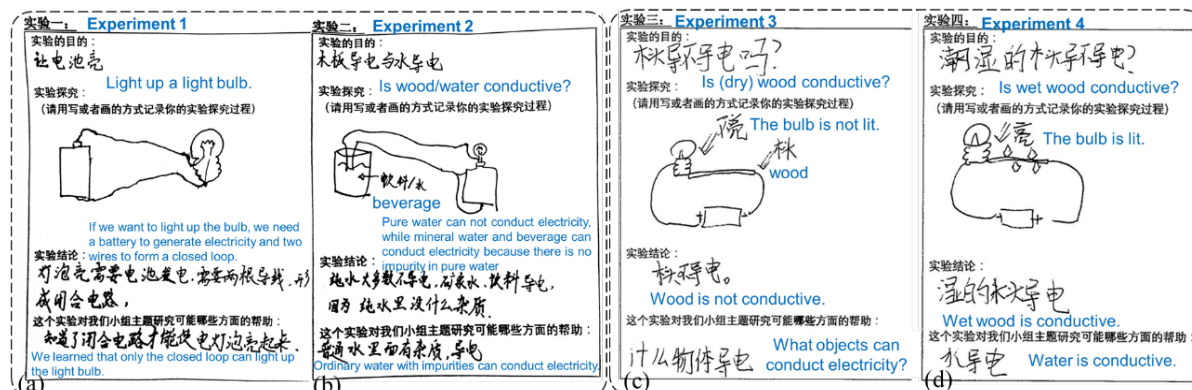


Figure 4. Experiment reports of Group 9 ((a) and (b)) and Group 5 ((c) and (d))

### View: Circuit

Then we focused on Group 5, who worked on *Circuit* in the experimental class. After they acquired the basics of circuit and successfully connected circuit to light up a light bulb, they began to explore deeper knowledge of electricity from other views. Therefore, they first identified the keywords “wood” and “dry” as idea friends from their group-level IFM in Week 5 (Figure 2 (b)), went to the *Conductor and Insulator* view to get the information they wanted, made a group plan for exploration, and did experiments to explore why wet wood can conduct electricity while dry wood cannot (Figure 4, (c) and (d)). Finally, Group 5 and Group 9 students collaborated to explain their common experimental phenomena and created a theory-building note (below) to explain the conductivity of different waters and wood.

[Putting our theories together] 1. Mineral water/beverage is conductive because there are ions that can move freely; pure water can't conduct electricity because there are no ions that can move freely. 2. Why wet wood is conductive: wet wood contains water, water is conductive.

The *Circuit* topic is the foundation of *Electricity*, so Group 5 students began to explore other important topics from Week 5, at which point the TDC value began to show a slight downward trend, as shown in Figure 3 (b); however, the TDC value in the experiment class remained higher than that of the comparison class. Students in the comparison class were concerned with fact-seeking questions—e.g., “What is voltage?”; “What is the unit of voltage?”; “What contribution does Ampere make to electricity?”; and “Where can I see the circuit in my life?”

### Class differences in the opportunistic community view

To characterize how community-level IFM helped students advance collective knowledge, we analyzed students' activities in the community view. First, we compared five categories of big problems identified in each class (Table 2). Results show that the experimental class identified more and deeper problems, while the comparison class tended to raise fact-seeking questions. This phenomenon is consistent with the notes coding results in RQ1.

In addition, students in both classes were invited to explore relationships, summarize, and create new knowledge according to the identified problems in Table 2. Students in the experimental class could conduct those activities smoothly with the support of community-level IFM (Figure 2, (c) and (d)), while students in the comparison class could only build on their original notes. We selected some community notes from the experimental class to show how those students synthesized, summarized, and created new knowledge through IFM.

### Theme 1: Synthesizing different topics

Student c705 from Group 5 found the keywords of the two topics “Franklin” and “metal” on IFM (Figure 2 (c)) were particularly close, then they tried to explain “why Franklin did not die” from the perspective of “why metals can conduct electricity”.

**[I want to discuss this topic]** Why did Benjamin Franklin not die from his lightning strike?

**[I want to explore relationships among those keywords]** metal, Franklin and copper key

**[My theory]** Thunderbolt voltages range from hundreds of millions to billions of volts, but Franklin was not electrocuted. We can conclude that he did not directly touch the copper key. Even if using Leiden bottles and ribbons, people will die if they touch a few hundred volts (unless the voltage is very, very small).

### Theme 2: Summarizing a key theme

Students c718, c726 and c711 from Group 6 found there were many keywords around “Franklin” in Figure 2 (c) and speculated those might be factors in protecting Franklin from the lightning strike. Then, they wrote down those keywords and selected relative information from the community.

**[I want to discuss this topic]** Why did Benjamin Franklin not die from his lightning strike?

**[I want to summarize those keywords]** Voltage, Leiden bottle, capacitor, wire, copper key, lightning rod, wind, ribbon, current, and metal.

**[Putting our theories together]** 1. Voltage (  *Low voltage* by student c751); 2. Leiden bottle (  *Low voltage* by student c722); 3. Iron wire (  *Iron wire* by student c723); 4. Copper key (  *Synthesizing: Franklin and Metal* by student c705); 5. Lightening rod (  *Lightening rod* by student c713); 6. Ribbon (  *Ribbon* by student c722); 7. Current (  *Current* by student c723).

### Theme 3: Creating new knowledge

Students c733 and c734 from Group 4 took responsibility for *Static Electricity* and found that the keywords “cloud” and “insulator” were very close in Figure 2 (c). Then they searched for information, identified related notes, and finally came up with new “knowledge”: “the cloud is an insulator.”

**[I want to discuss this topic]** What is the object can create static electricity?

**[I want to create new knowledge on these keywords]** cloud and friction.

**[My theory]** Our new knowledge is “the cloud is an insulator.” Evidence: 1. Electrostatics are all insulators (from Baidu); 2. Lightning is caused by frictions among clouds (from notes)

Students c708 from Group 5 and c752 from Group 2 provided more evidence to support the new knowledge—i.e., water evaporates into pure water vapour, which in turn forms clouds. Experiments proved pure water is an insulator, so the resulting cloud is also an insulator. Finally, we can see the new knowledge of “evaporation” in the next-week community-level IFM (Figure 2(d)). These data suggest that these young children, even in their simple writing, had strong knowledge-creation abilities, with the aid of IFM.

## Discussion

This study investigated how learning analytics can be used to promote knowledge building in a Grade 5 science classroom. First, we compared differences in learning outcomes, KF participation, and depth of knowledge building discourse in both experimental and comparison classes. Students in the experimental class showed more improvement in science concept understanding, higher social dynamics of reading, referencing and interacting across different subtopics, and a deeper discourse on theory building and synthesizing. Analysis of students’ group plans, experiment reports, and summary notes revealed how students used both group- and community-level IFM to identify learning gaps, make plans, co-construct, collect evidence, synthesize key ideas, summarize, and create new knowledge. Even though the IFMs are not concurrent, representations of learning analytics could be helpful suggesting some practical use in classrooms. Furthermore, we integrated learning analytics with social configurations in large classes—i.e., interactive groups with group-level IFM, and opportunistic interactions with community-level IFM—provide a pedagogical design for the less explored field of knowledge building in large communities (Scardamalia & Bereiter, 1996).

## References

- Baker, R. S., & Inventado, P. S. (2014). Educational Data Mining and Learning Analytics. In J. A. Larusson & B. White (Eds.), *Learning Analytics: From Research to Practice* (pp. 61-75). New York, NY: Springer New York.
- Burtis, J. (1998). The analytic toolkit. *The Ontario Institute for Studies in Education, The University of Toronto: Knowledge Building Research Team*.
- Chen, B., Scardamalia, M., & Bereiter, C. (2015). Advancing knowledge-building discourse through judgments of promising ideas. *International Journal of Computer-Supported Collaborative Learning, 10*(4), 345-366.
- Chuy, M., Zhang, J., Resendes, M., Scardamalia, M., & Bereiter, C. (2011). *Does contributing to a knowledge building dialogue lead to individual advancement of knowledge*. Paper presented at the Connecting computer-supported collaborative learning to policy and practice: CSCL2011 conference proceedings.
- Fu, E. L. F., van Aalst, J., & Chan, C. K. K. (2016). Toward a classification of discourse patterns in asynchronous online discussions. *International Journal of Computer-Supported Collaborative Learning, 11*(4), 441-478.
- Hakkarainen, K. (2003). Progressive inquiry in a computer-supported biology class. *Journal of Research in Science Teaching, 40*(10), 1072-1088.
- Kimmerle, J., Cress, U., & Hesse, F. W. (2007). An interactional perspective on group awareness: Alleviating the information-exchange dilemma (for everybody?). *International Journal of Human-Computer Studies, 65*(11), 899-910.
- Ma, L., Matsuzawa, Y., Chen, B., & Scardamalia, M. (2016). *Community knowledge, collective responsibility: The emergence of rotating leadership in three knowledge building communities*. Paper presented at the Proceedings of the 12th International Conference of the Learning Sciences, International Society of the Learning Sciences, Singapore.
- Matsuzawa, Y., Oshima, J., Oshima, R., & Sakai, S. (2012). Learners' use of SNA-based discourse analysis as a self-assessment tool for collaboration. *International Journal of Organisational Design and Engineering, 2*(4), 362-379.
- Nistor, N., & Hernández-García, Á. (2018). What types of data are used in learning analytics? An overview of six cases. *Computers in Human Behavior, 89*, 335-338.
- Oshima, J., Oshima, R., & Matsuzawa, Y. (2012). Knowledge Building Discourse Explorer: a social network analysis application for knowledge building discourse. *Educational technology research and development, 60*(5), 903-921.
- Oshima, J., Oshima, R., Murayama, I., Inagaki, S., Takenaka, M., Nakayama, H., & Yamaguchi, E. (2004). Design experiments in Japanese elementary science education with computer support for collaborative learning: hypothesis testing and collaborative construction. *International Journal of Science Education, 26*(10), 1199-1221.
- Resendes, M., Scardamalia, M., Bereiter, C., Chen, B., & Halewood, C. (2015). Group-level formative feedback and metadiscourse. *International Journal of Computer-Supported Collaborative Learning, 10*(3), 309-336.
- Ryan, J. (2013). *Education reform in China: Changing concepts, contexts and practices* (Vol. 69): Routledge.
- Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. *Liberal education in a knowledge society, 97*, 67-98.
- Scardamalia, M., & Bereiter, C. (1996). Engaging Students in a Knowledge Society. *Educational Leadership, 54*(3), 6-10.
- Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE review, 46*(5), 30.
- van Aalst, J. (2009). Distinguishing knowledge-sharing, knowledge-construction, and knowledge-creation discourses. *International Journal of Computer-Supported Collaborative Learning, 4*(3), 259-287.
- van Aalst, J., & Truong, M. S. (2011). Promoting knowledge creation discourse in an Asian primary five classroom: Results from an inquiry into life cycles. *International Journal of Science Education, 33*(4), 487-515.
- Wise, A. F., Vytasek, J. M., Hausknecht, S., & Zhao, Y. (2016). Developing Learning Analytics Design Knowledge in the "Middle Space": The Student Tuning Model and Align Design Framework for Learning Analytics Use. *Online Learning, 20*(2), 155-182.
- Zhang, J. W., Scardamalia, M., Reeve, R., & Messina, R. (2009). Designs for Collective Cognitive Responsibility in Knowledge-Building Communities. *Journal of the Learning Sciences, 18*(1), 7-44.

# Using and Perceiving Emoji in Design Peer Feedback

Na Sun, Pennsylvania State University, nzs162@psu.edu  
Elise Lavoué, University of Lyon, University Jean Moulin Lyon 3, iaelyon school of Management, CNRS,  
LIRIS, elise.lavoue@univ-lyon3.fr  
Chulakorn Aritajati, Pennsylvania State University, cya5092@psu.edu  
Aurélien Tabard, Université Lyon 1 & CNRS, aurelien.tabard@univ-lyon1.fr  
Mary Beth Rosson, Pennsylvania State University, mrosson@psu.edu

**Abstract:** Emoji are often used to express emotions or visually enrich text communications, but little is known about the impact of emoji in peer feedback. We conducted a qualitative study on emoji's roles in peer feedback in a small group collaborative design activity; the activity was conducted entirely online as part of a distance education class. Using content analysis on the feedback peers provided, and thematic analysis on transcribed interviews, we analyzed how emoji are used and perceived in four types of design feedback. Our analysis suggests that emoji increase peers' awareness of affective status, and are also used to mark minor cognitive feedback with a more personal flavor as part of document-based conversations. We found that emoji help to establish a general informality for peer feedback activities. We discuss the implications for the design of collaborative learning software that supports richer peer feedback.

## Introduction

As Värlander argues, “emotions should not be considered as hindering learning. Rather, it underlines the focal role of emotions in learning as being a natural part of it” (Värlander, 2008). Delivering and receiving feedback among peer learners is more than simply transferring information (Falchikov, 2013; Jacobs, 1974); each emotion-containing communication is influenced by students' emotions at the time of expression (Race, 1996). With the prevalence of emoji and their prospects of expressing emotions, this paper examines the use and perception of emoji in peer feedback.

## The role of emoji in peer feedback work

Emoji are a very popular manner of expressing emotions in real-time communication, and have also become increasingly prevalent in a range of social computing technologies (e.g. Twitter, Instagram, SNSs). In an educational context, Zhang et al. showed that students annotated discussion posts with emoji or tags to signal their confusions or curiosity; these signals directed instructors' attention to places where students needed help or further clarification (Zhang, Igo, Facciotti, & Karger, 2017). In this case, the use of emoji rather than text-based hashtags offered a relatively easy mechanism for conveying emotions. Such findings suggest that the use of emoji among students may allow for transparent emotional exchanges and further direct instructors' cognitive efforts. This motivated us to investigate the role of emoji in peer feedback within small groups in online classes as a way to facilitate social interaction among distance learners.

One recent work leveraged crowdsourcing intelligence to express emotions in feedback with pictures versus text (Robb, Padilla, Kalkreuter, & Chantler, 2015). The designers who created the work being critiqued felt that both text and emotional images were more “threatening” than an abstract image; they also reported that the abstract images were inspiring. Nonetheless, text-based feedback is more likely to be used for giving conventional critiques: In a follow up study that examined how crowd workers felt when giving feedback using abstract images, emotion images and text, Robb et al. revealed that the type of feedback preferred was correlated with users' cognitive styles, and that over half of the participants valued engagement over clarity (Robb, Padilla, Methven, Kalkreuter, & Chantler, 2017). Furthermore, some participants who provided feedback said that they would have liked to use emoji in addition to image-based emotion feedback. Although images have been found more expressive for emotions, some users find ambiguity to be a desirable advantage of giving emotion feedback because it is harder to put emotions in words than to use an image, and emoji are prone to ambiguity due to varying interpretations (Miller et al., 2016).

In addition to expressing emotions for the user, the use of ambiguous and playful expressions of emotions may invite and leave more space for communicative moves from the recipients, which could encourage more social interaction. In this regard, ambiguous and abstract indicators (e.g. emoji) seem consistent with Jakobson's model of communication (Jakobson, n.d.). In that model, the simple act of continuing a



conversation is valuable. Indeed, emoji have been used to manage conversation (Pohl, Domin, & Rohs, 2017), and even have been perceived as a more efficient and playful way of expression than text.

As a complement to text-based comments, emoji may provide additional emotional and situational information to the recipients. For example, a study that examined senders' intentions in US-based messaging reported that emoji are used as a social tool to engage the recipients in a conversation, to convey intended meaning, and to solicit certain reaction from the recipients, even when the senders intended nothing specific in the emoji themselves (Cramer, de Juan, & Tetreault, 2016). More generally, in a quantitative study of Twitter data (Pohl et al., 2017), emoji have been used as message decoration, to replace words, and to enrich text messages with respect to tones or meaning, in addition to serving the function of emotional annotation.

Despite the fact that emoji have become prevalent as part of a growing and universal language in text-based communication, researchers have not yet understood the role of emoji in remote collaborative learning tasks. Emoji seems to be a promising but under explored option for engaging online learners in conversations with additional situational information. We are especially interested in its potential in expressing emotional information in online learning environments where students often feel isolated and yearn for social connections with others (Sun, Rosson, & Carroll, 2018). However, emoji are subject to misinterpretations according to individual differences and varied rendering mechanisms across devices (Miller et al., 2016). Although it is possible that reacting to others' learning artefacts may stem from different interpretation and lead to differing emotional states, our study focuses on the impact of such emotions in interactive moves and interpersonal relationships situated in social learning activities. We now turn to our study of emoji as a mechanism for adding emotional content to peer-based critiques as part of a collaborative learning activity.

## Research setting












Situated in collaborative design activities among group members of an online bachelor degree program, we designed an online feedback tool called Emoviz to probe how peer feedback with emoji-entries affects the social interactions and relationships among online students.

## Selecting appropriate emoji for emotion feedback

Previous studies found that emoji can be perceived very differently within and across platforms, and thus lead to senders' and receivers' misunderstandings (Tigwell & Flatla, 2016). In addition, there are certain emoji with complex emotion interpretations that can cause high levels of sentiment misconstrual. There seems to be a tradeoff between communicative expressiveness and the consistency of emoji interpretation. To wit, the simpler an emoji is, the more uniform are people's interpretations (Miller et al., 2016). To make our studies more generalizable and comparable with existing studies, we adopted the Unicode standardized emoji, displaying them consistently as a Portable Network Graphics (png) file to minimize the cross-device differences.

We used existing emotion frameworks and emotion-related learning models (D'Mello, Lehman, Pekrun, & Graesser, 2014; Pekrun, Goetz, Titz, & Perry, 2002; Plutchik, 2003) to we select eleven emotions that are relevant to academic peer feedback and design activity. These emotions include *joy*, *surprise*, *praise*, *pride*, *love*, *anger*, *confusion*, *anxiety*, *disapproval* and *boredom*. Considering the functions of emoji in social media research, we further included *playfulness* to extend the corpora of emotions in a broader sense (Pohl et al., 2017). As Table 1 shows, we chose three emoji as representatives for each emotion and proceeded to investigate the effectiveness of each one with a survey distributed to crowdworkers.

Table 1: Selecting representative emoji for relevant emotions for feedback

|          |   |            |             |  |            |
|----------|---|------------|-------------|--|------------|
| joy      |  | n.s.       | confusion   |  | $p < .001$ |
| surprise |  | n.s.       | anxiety     |  | n.s.       |
| praise   |  | $p < .001$ | disapproval |  | $p < .001$ |
| pride    |  | n.s.       | boredom     |  | $p < .001$ |
| love     |  | $p = .001$ | playfulness |  | $p < .001$ |
| anger    |  | $p < .001$ |             |  |            |

We obtained 72 responses from Amazon Turk where we asked MTurkers to rate "To what extent do you agree that the following emoji represent <selected emotion>," on a Likert scale from 0 (Strongly disagree) to 100 (Strongly agree), excluding three that failed the quality test questions (we inserted two questions with



clear answers to select the qualified respondents for emoji interpretations for our survey). Our participants are all Americans (n = 69, 36 females), aged between 19 to 65, at an average age of 35.48. A majority of them (88.4%) used emoji in their online communications at least sometimes. Based on the General Linear Model to compare repeated measures, we found significantly different emoji for all emotions except for *joy*, *surprise*, *pride*, *anxiety* at the significance level of .001 with a Greenhouse-Geisser correction (see Table 1, highest emoji for each emotion is underscored). Regardless of the significance, the underlined emoji in Table 1 indicates which emoji candidates of each group were rated highest in the triplet. After this process, we selected the highest-crowd-rated emoji to represent each emotion, except for the confusion emotion; in that case we chose the higher-rated emoji that was also graphically consistent with the other set (i.e., 😞 instead of ?).

## Emoviz: An emoji-enabled online annotation tool

Emoviz is the online feedback tool depicted in Figure 1, offering customized regions for comments in text and/or emoji. Users can drag a rectangle over anywhere in a PDF file to leave comments (see Figure 1); the emoji included in the software were carefully selected through the aforesaid crowdsourcing study. They are displayed against a light-yellow rectangular background; any emoji can be appended to the end of a text comment. We adapted Emoviz for within-group feedback activities: students can upload documents in PDF format, give and receive peer feedback directly on the document. In addition, when comments are generated in the PDF document, a notification is sent to the PDF owner every 24 hours so that the submitters are aware of the feedback made to their document.

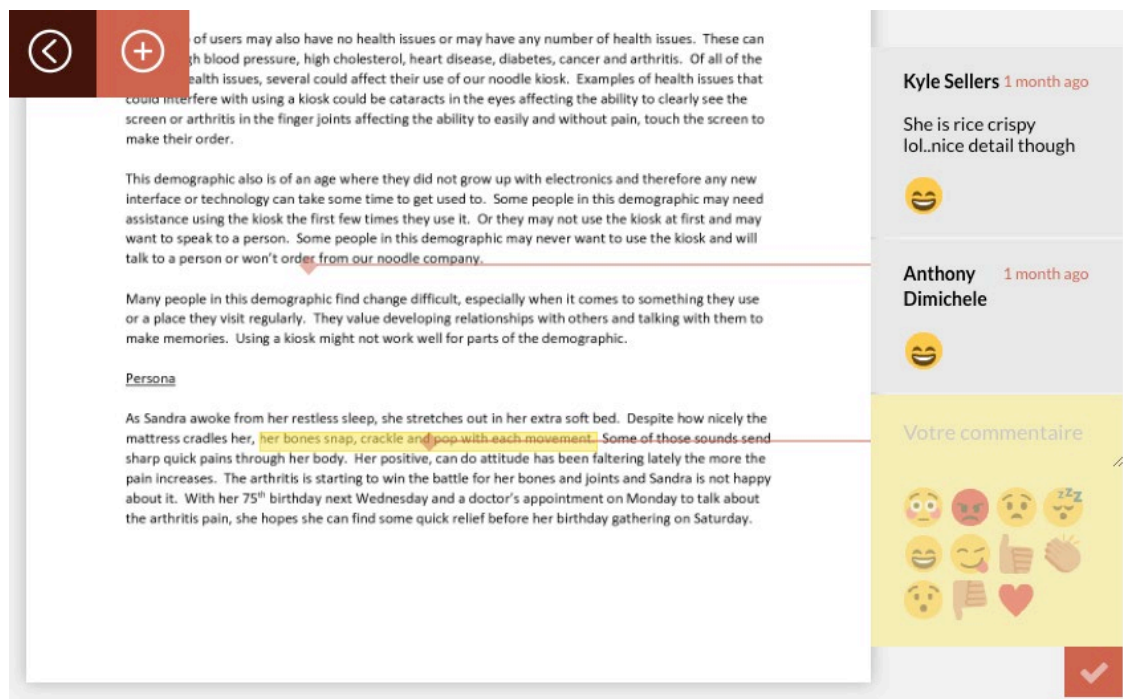


Figure 1. Emoviz User Interface for Giving and Receiving Feedback on PDF documents.

## Method

### Procedure

Our study took place in an User-centered Design course offered by an online bachelor degree program of a Northeastern American university. Across the semester, the 50 students are tasked with a mix of individual assignments and group projects. In particular, our study built on two individual assignments that were embedded within a larger group project: one is the persona assignment, which is individual work but supposed to have other group members' feedback; another is the low fidelity prototype turned in PDF by individuals, aiming to yield a final prototype that the whole group agrees to proceed with (either building on top of one individual's submission or collaboratively generating a new prototype with all individuals' input). At the 8th week of the Fall 2017, we recruited students for the aforesaid two design activities via an opt-in invitation: all students of the

class were invited via Email link sent from Emoviz system, preceded by a call for participation sent by the course instructor on behalf of the research team and followed by a post-survey.

## Data collection and analysis

Throughout this deployment of Emoviz, we collected both quantitative and qualitative data and analyzed the emoji use and feedback activities among our participants to triangulate our findings (see Table 2). As for the quantitative data, we collected survey responses after the first design activity with regards to their demographic information, group cohesiveness, frequency of using emoji and emotional awareness of other members in general. To measure group cohesiveness, we employed six items from a well-established scale (Podsakoff, Niehoff, MacKenzie, & Williams, 1993). Only 7 students out of the entire class filled in this survey, and group cohesiveness range from 5.29 to 6.29 on a 7-point Likert scale.

Table 2: Triangulated data for Emoviz use and feedback activity

| Data source | #  |
|-------------|----|
| Comment     | 51 |
| Survey      | 7  |
| Interview   | 6  |

To gather qualitative data, we recorded Emoviz use with respect to the uploaded design documents and comments left during the collaborative process. The first author applied open coding and axial coding to analyze the feedback content at the unit of a comment (n = 51) by first categorizing feedback content learners have left in the document, and then scrutinizing the emoji's role in different types of feedback. Specifically, we began with an existing feedback model that distinguish cognitive feedback (e.g. summarization, specificity of problems, explanations) and affective feedback (e.g. praise, mitigation) (Nelson & Schunn, 2009); in parallel we adopted elements of grounded theory to label feedback with emergent codes (Strauss & Corbin, 1998).

In addition to classifying feedback content and form, we interviewed six Emoviz users remotely over Google Hangout after the design activities had been completed. We chose to use Google Hangout because this particular real-time communication tool is quite familiar to online students (Sun & Rosson, 2017). The interviews lasted 53 minutes on average, and were transcribed by the first author. Using MAXQDA12 software, the first author generated initial codes and reviewed and iterated the codes on using thematic analysis (Braun & Clarke, 2006). In particular, we found these themes: general emotional connections felt in the remote group; task reflection; other tools they have used for group activities in the class; usability issues of Emoviz; barriers they faced in using emoji or expressing emotions with other peers; and the effect of using and interpreting emoji. For the purposes of this paper, we only present the codes from interview related to the use of emoji and the perceived functions of using emoji in peer feedback activities.

## Participants

For our study, 13 students uploaded at least one design assignment for peer feedback activities from six groups in Emoviz. However, only 10 students representing three different groups received comments from other members, while the other three students' uploaded design work was unmarked. Among the participants, two were female, and half were Caucasian, followed by African American and Mixed Race. All but one student were working full-time while pursuing online degrees. Most students were in their fourth year of study (mean age = 33.14) and thus can be viewed as relatively experienced online learners.

## Results

Among the 13 documents that received comments from peers, we found 51 comments from eight different learners; 22 of these contained a total of 30 emoji. Based on an existing cognitive and affective feedback model (Nelson & Schunn, 2009) and open-coding of feedback that was not covered by that model, we identified four types of feedback towards the design work or described scenarios in Emoviz (note that these are not mutually exclusive, see table 3): 1) cognitive feedback (e.g. specific suggestions); 2) affective feedback (i.e. positive or negative); 3) conversational feedback (e.g. raising attention, showing idiosyncrasy, and responses to others' comments); 4) comparative feedback based on feedback givers' self-reflections. Accordingly, by a close examination on each comment and the context in which the comment is made, we identified different roles emoji play in these different types of feedback situations. In the following sections, we use the notation that

combines the data collection name and the participant ID in the following sections. For example, IP28 refers to participant 28 in the interview data, whereas DP28 refers to the same person's data in the content analysis.

Table 3: 4 Types of feedback and the use of emoji

| Type          | Cognitive   | Affective  | Conversational   | Comparative  |
|---------------|---|--|--|--|
| # of comments | 19  | 43   | 20   | 10   |
| # of emoji    | 0   | 28   | 18   | 7  |
| Examples      | DP13:<br>“You probably want<br>“the” instead of “to”<br>here” | P13: “Clean looking<br>recap screen option!<br>👍”<br>DP13: “Great size<br>option! 😊❤️” | DP38: “She is rice<br>crispy lol...nice<br>detail though 😊”<br>DP27: 😊 | DP33: “Well written<br>short story. So much<br>more interesting<br>than the bio I wrote<br>:)” |

### Emoji were used only with non-negative feedback

Emoji were more pervasively used in the comments that expressed favorable review of design elements; in contrast criticisms or suggestions for improvement were shared as text only. More specifically, we observed that all the cognitive feedback (e.g. suggestions for improvement or critiques) was articulated in a text-only manner. However, participants wrote 28 positive comments to recognize, compliment or make jokes about others’ work, and the remaining 23 comments were either neutral or negative, for example being used to ask for clarification, point out errors and to provide specific suggestions for improvement.

26 out of the 30 emoji and emoticons (e.g. :-)) or :)), such as 👍, 😊, were used to praise or acknowledge others’ good work, for example commenting on a clean layout or a specific design element. Although the numbers are too small to make a statistical comparison, there seemed to be a tendency to prefer emoji when commenting on graphical elements, such as persona photos or design sketches. For example, DP21 commented on a graphic photo at corresponds to DP28’s persona description, “*What a serious looking man. Doesn’t look like he’s actually smiling 😊😏👍*,” whereas DP28 responded with a 😊👍.

Only two comments leveraged emoji for negative emotions (i.e. 🙄, which, instead of being a critique to the writing or design work, was a reaction to the story characters’ unhealthy habits or behaviors. For example, DP40 commented “*Whoa! 😱*” in response to DP33’s persona description “*Bob ordered nine cheese burger*”. In general, 👍 were used most often (10 times), followed by 😊 (8 times), and 😏 (4 times); meanwhile, most of negative emoji were never used (e.g. 😡, 🙄, 😞).

Our follow-up interviews revealed that students chose not to use emoji in negative comments because the emoji available for negative emotion (e.g., 🙄), were perceived as too aggressive for criticism that they wished to express in a mild way. For example, IP27, a stay-at-home Mom with 5 kids, would never use a sleepy face “*with people you don’t truly know that well- you only know them through group experience.*” IP25, a senior male student, also felt it more appropriate to use positive emoji, thinking that “*the thumbs down could be a bit aggressive, especially if you are trying to approach from a ‘yes, and’ perspective and not the ‘yes, but’ one.*”

Indeed, honest expressions of critiques were also associated with perceived closeness among collaborators: IP38 suggested that he feels comfortable joking with familiar teammates, those who he regarded as “*friends*” even without emotional tones, such as “*lol,*” as opposed to “*when you are saying a statement you want to make sure the receivers understand you are not being serious what you are saying*”. The unfamiliarity and lack of intimacy prevent students of a virtual group from giving negative comments more frankly.

### Signaling insignificant cognitive feedback with emoji markers

A closer scrutiny on the 22 comments that contain emoji revealed that, aside from the association between emoji and positivity, these comments were often informal, casual, and light-hearted – anything but serious call for attention. Examples include DP 38’s comments on the described character, such as “*Ha! They are lazy 😊*”, and DP13’s reaction to the drawings of four types of bread in a User Interface sketch: “*Oprah ‘I Love Bread’ and so do I 😊😏*”. Such comments seemed to contain a hint by the comment giver that a comment is meant to be expressing emotions at the moment (e.g. cheerful, surprised, etc.) IP33 appreciated the expressiveness of emoji, commenting “*without emoji it is just hard to convey what you are feeling about something - it is a very simple (way) of doing that.*” IP40 liked how emoji can articulate her real reaction in response to the described scenario

at the moment: Looking at the document she commented on, she said, “That emoji face I put there was exactly the big sad face I had after I read it (giggling). It is like a ‘Wow’. You know a jaw drop reaction.” She liked how the facial expressions of emoji in the feedback represented her real reaction.

IP34 further noted that emoji of different facial expressions enabled more nuanced expression and capture of feedback givers’ emotional statuses: “(if) you have somebody laugh with tears flowing from it- I tell you there is a person who is cracking up or the person is really amused compared with a mild amusement. Yeah, so emoji do make differences in those occasions.” He further explained that the explicit emotional annotations with emoji denoted one’s real social indication better than audio, “if a person uses emoji, you precisely get that indication, but other than that, (like) audio chat without visuals, it is almost as limited as just text without any emoji.” In this regard, emoji provide subtle social cues that enlighten the social interaction in an online environment where emotional statuses are hard to convey or perceive.

In addition to articulating the emotions right at the moment, emoji also helped to provoke an interesting conversation, by communicating an open posture of being engaged or sociable. For instance, 12 out of 20 conversational comments were decorated with emoji. IP34 attributed emoji to cues that remote collaborators “can easily pick and follow,” since use of emoji indicates “when somebody is relaxed and engaged when we are having those conversation compared with someone (who) is not really engaged.” Therefore, use of emoji gives students like IP34 a sense of “if somebody is open to more communication.” IP40 suggested that emotional elements evoked through humor also help online student get through course projects easier as they come and go from one course to another. For instance, after reading DP 27’s description about an old grandma user “her bones snap, crackle and pop with each movement” (Figure 1), DP38 joked about how the character is extremely vulnerable, pointing out “she is rice crispy lol...nice detail though 😊,” and another team member DP13 responded to the joking comment with a 😊. DP38 explained in the follow-up interview that he liked to leave the impression of himself being humorous, and that emoji or its alternative, such as “lol”, allowed him to express his personality more boldly. IP32 also found emoji help to create feelings of closeness to their peers who used them. As another example, DP31 joked about the looking of the persona picture along with the described scenario in text format when the persona was hesitating of quitting his job: “Judging by the look of his picture, he plans to quit his job 😊”

Whether emoji are used to express emotions or initiate an interesting conversation, comments that contain emoji should not be taken seriously (as opposed to how to improve the design work or to praise). Such non-serious signals are also perceived and accepted by feedback receivers - students interpreted the traditionally “negative emoji” in a positive way, appreciating the reaction from the readers. For example, IP33 regarded 👍, 🙄 and 😬 altogether as positive feedback, articulating that he likes them. As he elaborated, “it shows that she read it and she was not really disagreeing but empathizing with what I have written.” In this way, emoji help to assure that designers’ creative composition, such as illustrative depiction of a character or situation, in the design work has been understood in the intended way. IP27 also found it a relief to see others’ funny comment on her carefully composed design persona, “the fact they talked about how they enjoyed, the funny comments in return saying that they enjoyed it makes me think that ‘okay, good’. It gave me a good feel as to whether or not what I was trying to say came across right, or what I said is alright to say.” This use of emoji extended how people usually use emoji to complement a text and express their emotions (Zhou, Hentschel, & Kumar, 2017) with enriched semantics of feedback content.

## Self-Reflection and Conversational Feedback

Surprisingly, we found a number of feedback comments that referred to a feedback-giver’s own work or situation. For example, DP33 wrote “Well written short story. So much more interesting than the bio I wrote :)” in DP40’s submission as a compliment while comparing it with his own submission. Description in a persona also sometimes evokes a sense of self-reflection, for example, DP27 felt a desirable state in reflection on her own tight schedule when reading DP13’s work, commenting that “Nice job! I kinda envy Noah’s availability. :-)” In another example, after reading the isolated living atmosphere of a grandma described in a scenario, DP38 expressed his empathy with the written character by DP27 in a sense of humor with “Thanks for the guilt trip 😬.” It seems the reviewing experience is not only a critique session about a peer’s work quality, but also sometimes a venue to share personal reflections with reference to personal work or experiences.

Meanwhile, we coded 20 comments as conversational feedback because of their dialogic nature, as if the feedback givers are talking with the feedback receivers in a social chit-chat: For instance, DP27 commented on the scenarios with an obvious conversational tone: “This is so very true, I empathize with this user! For sure!”. The emotional sentiments aroused by the design were depicted via emoji; Vice versa, as a feedback

receiver, IP27 found comments with emoji more impressive and fun to be remembered. In particular, she took content related comment as indicators that her original meaning has been clarified enough to be understood by the reader to the effect that is intended by the designer.

## Discussion

In contrast to previous studies that looked at emoji in message-based communication or social media (Miller et al., 2016; Zhou et al., 2017), our study is the first to examine the role of emoji in a group learning activity. Our study confirms the speculation from previous studies on emoji' advantage of representing ambiguous emotion feedback content (Robb et al., 2017). We describe how learners utilized the toning-up effects of emoji to signal non-verbal cues and personality, such as sense of humor and facial expressions. Embedding emoji as part of feedback options allows for more casual and in-situ marks of one's own emotion, and thus enhances perceived sociability in CSCL environments (Kreijns, Kirschner, Jochems, & van Buuren, 2007).

We also found that emoji served to initiate and manage conversation as part of a learning activity, as documented in social media contexts (Cramer et al., 2016). In this sense, emoji seemed also to carry over the norm of being socializing with other users as in social media platforms. Emoji in the peer feedback scenarios are used by feedback givers for initiating informal and light-weighted conversations. They are interpreted by feedback receivers as a social posture of the other interlocutor being friendly and responsive to their creativity and personality. In addition, we also observed the role of emoji to in eliciting empathized self-assessment and generating feedback based on social comparison with one's own design work. Therefore, our study points to the design opportunities of leveraging emoji to add more nuanced peer feedback content, such as depiction of in-situ emotional responses, and to elicit personal interaction among learners.

Further, we propose emoji can enrich the markers to specify the anchor points during the learning process beyond using emotional vs non-emotional markers (Lavoué, Molinari, Prié, & Khezami, 2015). With the lightened tones of emoji alongside the text comments, feedback givers can mark the "insignificance" of the feedback, enabling more nuanced layers of feedback types in addition to affective and cognitive feedback. In particular, feedback givers' use of emoji hints at the informal nature of the particular comment, and thus helps feedback receivers' distinguish casual content for entertainment, fun or other engaging gestures for social purpose from serious comments, which aligned with the conventional critique formats (Robb et al., 2015).

## Limitation and future work

Although we have found evidence of using emoji to support peer feedback interaction, we admit the potential limitation introduced by our small sample size and the particular online class environment where group members are relatively cohesive. Although we do not have enough data to investigate this point, the use of emoji may also encourage more casual discussions which may hinder knowledge building dynamics. Besides, our interviews asked participants to recall their emoji use and response after finishing the peer feedback task, which may lead to memory lapse and inaccurate descriptions.

As discussed in the literature, the interpretation of emoji depends on platforms and situational factors. For the purpose of the study, we controlled the number and design of emoji to learning related emotions. Future research could expand to open-ended use of emoji and study their relationship to learners' perceived emotions. More quantitative research could also examine the association between emoji use and feedback perceptions.

## Conclusion

Our study presents the first attempt of involving emoji in the context of design feedback activity, and also demonstrates ways of engaging distance students with emoji as feedback options in an asynchronous peer learning activity. Our findings revealed how emoji are associated with different types of feedback, and indicated the potential of emoji to transform a formal peer critique task into a more emotion-aware collaborative process. We also provide implications of utilizing informal language cues, such as emoji, in designing learning systems that consider and support socioemotional dimensions of online interaction.

## Acknowledgement

Our research was conducted as part of the EmoViz project funded by the Région Auvergne-Rhône-Alpes. We thank our student participants and Emily Wenk, the instructor who helped us recruit students.

## References

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.

- Cramer, H., de Juan, P., & Tetreault, J. (2016). Sender-intended Functions of Emojis in US Messaging. *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*, 504–509. <http://doi.org/10.1145/2935334.2935370>
- D’Mello, S., Lehman, B., Pekrun, R., & Graesser, A. (2014). Confusion can be beneficial for learning. *Learning and Instruction*, 29, 153–170. <http://doi.org/10.1016/j.learninstruc.2012.05.003>
- Falchikov, N. (2013). *Improving assessment through student involvement: Practical solutions for aiding learning in higher and further education*. Routledge.
- Jacobs, A. (1974). The use of feedback in groups. *The Group as Agent of Change*, 408–448.
- Jakobson, R. (n.d.). 1960. "Closing Statement: Linguistics and Poetics.". *Style in Language*, 350–377.
- Kreijns, K., Kirschner, P. a., Jochems, W., & van Buuren, H. (2007). Measuring perceived sociability of computer-supported collaborative learning environments. *Computers & Education*, 49(2), 176–192. <http://doi.org/10.1016/j.compedu.2005.05.004>
- Lavoué, É., Molinari, G., Prié, Y., & Khezami, S. (2015). Reflection-in-action markers for reflection-on-action in computer-supported collaborative learning settings. *Computers and Education*, 88, 129–142. <http://doi.org/10.1016/j.compedu.2015.05.001>
- Miller, H., Thebault-Spieker, J., Chang, S., Johnson, I., Terveen, L., & Hecht, B. (2016). “Blissfully happy” or “ready to fight”: Varying Interpretations of Emoji. *Association for the Advancement of Artificial Intelligence*, 259–268. <http://doi.org/10.1089/cyber.2011.0179>
- Nelson, M. M., & Schunn, C. D. (2009). The nature of feedback: How different types of peer feedback affect writing performance. *Instructional Science*, 37(4), 375–401. <http://doi.org/10.1007/s11251-008-9053-x>
- Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic Emotions in Students’ Self-Regulated Learning and Achievement: A Program of Qualitative and Quantitative Research. *Educational Psychologist*, 37(2), 91–106. <http://doi.org/10.1207/S15326985EP3702>
- Plutchik, R. (2003). *Emotions and life: Perspectives from psychology, biology, and evolution*. American Psychological Association.
- Podsakoff, P. M., Niehoff, B. P., MacKenzie, S. B., & Williams, M. L. (1993). Do substitutes for leadership really substitute for leadership? An empirical examination of Kerr and Jermier’ s situational leadership model. *Organizational Behavior and Human Decision Processes*, 54(1), 1–44.
- Pohl, H., Domin, C., & Rohs, M. (2017). Beyond Just Text: Semantic Emoji Similarity Modeling to Support Expressive Communication???. *ACM Transactions on Computer-Human Interaction*, 24(1), 1–42. <http://doi.org/10.1145/3039685>
- Race, P. (1996). The art of assessing 2. *NEW ACADEMIC-BIRMINGHAM-*, 5, 3–6.
- Robb, D. A., Padilla, S., Kalkreuter, B., & Chantler, M. J. (2015). Moodsource: Enabling Perceptual and Emotional Feedback from Crowds. *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing*, 21–24.
- Robb, D. A., Padilla, S., Methven, T. S., Kalkreuter, B., & Chantler, M. J. (2017). Image-based Emotion Feedback. In *Proceedings of the 2017 Conference on Designing Interactive Systems - DIS '17* (pp. 451–463). <http://doi.org/10.1145/3064663.3064665>
- Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Sage Publications, Inc.
- Sun, N., & Rosson, M. B. (2017). Finding the Community in Online Education: It’s in the Instructors’ Eyes. Philadelphia, PA: International Society of the Learning Sciences.
- Sun, N., Rosson, M. B., & Carroll, J. M. (2018). Where is Community Among Online Learners?: Identity, Efficacy and Personal Ties. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (p. 292). ACM.
- Tigwell, G. W., & Flatla, D. R. (2016). “ Oh that ’ s what you meant !” : Reducing Emoji Misunderstanding. *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct (MobileHCI '16)*, (Cmc), 859–866. <http://doi.org/2957265.2961844>
- Värlander, S. (2008). The role of students’ emotions in formal feedback situations. *Teaching in Higher Education*, 13(2), 145–156. <http://doi.org/10.1080/13562510801923195>
- Zhang, A. X., Igo, M., Facciotti, M., & Karger, D. (2017). Using Student Annotated Hashtags and Emojis to Collect Nuanced Affective States. *Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale - L@S '17*, 319–322. <http://doi.org/10.1145/3051457.3054014>
- Zhou, R., Hentschel, J., & Kumar, N. (2017). Goodbye text, hello emoji: mobile communication on wechat in China. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 748–759). ACM.



# On the Understanding of Students' Learning and Perceptions of Technology Integration in Low- and High-Embodied Group Learning

Marianna Ioannou, Cyprus University of Technology, marianna@cyprusinteractionlab.com  
Yiannis Georgiou, Cyprus University of Technology, Research Center on Interactive Media, Smart Systems and Emerging Technologies (RISE), yiannis@cyprusinteractionlab.com  
Andri Ioannou, Cyprus University of Technology Research Center on Interactive Media, Smart Systems and Emerging Technologies (RISE), andri@cyprusinteractionlab.com  
Mina Johnson-Glenberg, Embodied Games, Arizona State University, minaj@embodied-games.com

**Abstract:** Embodied learning activities supported by motion-based technologies are becoming popular in various contexts and settings. However, little is yet known about technology integration for embodied learning in groups in authentic classroom settings, as existing studies have been mostly conducted in laboratory settings. In this work, we examine students' learning and perceptions of technology integration for group learning in a high-embodied, Kinect-based educational game (Condition1, n=24 students), in comparison with a low-embodied, desktop-based version of the same game (Condition2, n=18 students), in an authentic classroom setting. Data collection included questionnaires evaluating students' baseline, knowledge gains, perceptions of technology integration, and post-activity interviews. Findings showed higher learning gains and more positive perceptions of technology integration, in terms of opportunities for student negotiation, for the students in the low-embodied condition. Implications are discussed for supporting high-embodied learning activities for group learning in authentic classroom settings.

## Introduction and theoretical background

Embodied learning, as an application of the embodied cognition theory, constitutes a contemporary pedagogy of learning, which emphasizes the use of the body in the educational practice (Antle, 2013, 2009; Barsalou, 2010; Georgiou & Ioannou, in press). Embodied learning environments based on motion-based technologies appear to have gained ground during the last decade, providing new ways in which the embodied cognition theory can be enacted in the field of education (Kosmas, Ioannou, Retalis, 2018, 2017). In particular, the widespread population of affordable motion-based technologies and natural user interfaces (e.g., Wii, Xbox Kinect, Leap Motion) have nowadays opened the doors for the design of technology-enhanced embodied learning environments, responding to the need to highlight the aspects of motion and physicality as a crucial part of the learning process (Melcer & Isbister, 2016). As argued by Maliverni and Pares (2014), technology-enhanced embodied learning environments open up new possibilities due to their potential affordances for promoting psychomotor, cognitive and affective learning gains.

From a technology integration perspective, incorporating innovative motion-based technologies and natural user interfaces in the school classroom introduces new challenges to teaching and learning; therefore, their incorporation in mainstream education is at a very slow pace (Abrahamson & Sánchez-García, 2016). According to Karakostas, Palaigeorgiou, and Kompatsiaris (2017) existing research on embodied learning technologies has been fragmented and is driven largely by technological innovations. Also, it has taken place mostly in laboratory settings focusing on the participants' interactions with the embodied environments, therefore lacking a clear focus on investigating their efficacy in authentic educational contexts (Karakostas et al., 2017). Moreover, while a number of technology-enhanced embodied learning environments have received positive evaluation in high-controlled laboratory settings (e.g., Homer et al., 2014; Lindgren, Tscholl, Wang, & Johnson, 2016), the limited number of studies conducted in authentic school classrooms had not be as successful as initially expected in promoting students' learning gains, compared to low-embodied, desktop-based environments (e.g., Anderson & Wall, 2016; Hung, Lin, Fang & Chen, 2014). The later warrants further investigation towards shedding light on the efficacy of high-embodied vs. low-embodied learning environments in authentic classroom settings.

The study adopts Moos (1987)'s conceptual framework of technology integration grounding the experience in three dimensions: "Relationship", "Personal development" and "System maintenance and change". Moos (1987)'s framework was used by Wu, Chang, and Guo (2007) and Maor and Fraser (2005) to derive subscales for the measurement and evaluation of the experience of technology integration. In particular, in this work, we examine students' perceptions of technology integration aimed at a high-embodied, Kinect-based learning experience (Condition1, n=24 students), in comparison with a low-embodied, desktop-based learning

experience (Condition2, n=18 students) for leaning in groups in an authentic classroom setting. The study sought to answer the following research questions: (i) *Are there differences in student's learning gains and attitudes between the conditions?* (ii) *What are the main factors affecting student's perceptions of technology integration in the two conditions?*

## Methodology

### Participants

Participants were 42 students in 4<sup>th</sup> grade (aged 8-9 years old), who were enrolled in a public primary school in the Eastern Mediterranean. The students were randomly assigned to the two conditions. Group1 (Kinect-based gaming condition) had 24 children (12 boys, 50%) and Group2 (Desktop-based gaming condition) had 18 children (11 boys, 61%).

### Research design

This study followed an explanatory sequential design, composed of two sequential phases (Creswell, Clark & Vicci, 2011). In phase 1, we adopted a two-group quasi-experimental design for investigating students' learning gains and perceptions of technology integration in both conditions. In phase 2, we proceeded with qualitative data collection to deepen our understanding of the factors relating to students' perceptions of technology integration.

### The digital game

We employed the "Alien Health" digital game, which was designed to teach 4<sup>th</sup>-12<sup>th</sup> grades about nutrition and healthy food choices. The game is well-related to the school curriculum whilst findings from previous studies of "Alien Health" indicated its acceptability by the children and its affordances to improve content knowledge (Johnson-Glenberg & Hekler, 2013). Children's mission in the game is to make the right nutritional choices for the alien to make him feel better as he is in charge of stopping the collision of an asteroid with the Earth. During the gameplay, children are presented with combinations of food and are requested to make choices within predefined timeframes, considering a constellation of five nutrients per food (protein, fats, carbohydrates, fiber, and vitamins/minerals). The digital game became available in both a low-embodied (desktop-based) and in a high-embodied (Kinect-based) version.

### The interventions

Considering the research goals of this study, an 80-minute intervention was developed for each condition. Children in the low-embodied (desktop) condition were divided in dyads and used the desktop-based version of the digital game. In this version children used the mouse and the keyboard for making a choice and feeding the alien (see Figure 1). Children in the high-embodied (Kinect-based) condition were divided in groups of four (the limited classroom space allowed only six Kinect work-stations of the game) and used the Kinect-based version of the game. In this case, the game was projected on a big screen and there was touchless interaction via the Kinect camera which can identify children's arm/hand movement hovering over a single food item and moving it into the Alien's mouth (see Figure 2). In both conditions the game was contextualized in a collaborative educational activity. In particular, the children took turns in playing (game affords only single player mode); the other child(ren) of the group was/were asked to provide feedback to the player, discuss the selections made, and record their food choices on a structured worksheet.



**Figure 1.** A group of students in the low-embodied condition.



**Figure 2.** Two group of students in the high-embodied condition.

### Data collection and analysis

Data collection included questionnaires evaluating students' baseline, knowledge gains, perceptions of technology integration, and post-activity interviews with eight students from Condition1 (33.3%) and eight students from Condition2 (44.4%).



### Baseline data

We aimed at establishing the equivalency of the two conditions in terms of computer and gaming attitudes. We used the Computer Attitude Measure for Young Students questionnaire (CAMYS, Teo & Noyes, 2008), which is composed of 12 items on a five-point Likert scale and has a documented reliability alpha coefficient of .85. Gaming attitudes were measured using an 11-item Likert scale (Cronbach's alpha=0.73) validated in the study of Bressler and Bodzin (2013). Differences between the two conditions were examined using the Mann-Whitney U test, given the small sample size of participants in each condition and the lack of normal distribution in the data.

### Knowledge test

A knowledge test was administered in pre-post format. The test was developed by Johnson-Glenberg and Hekler (2013) for evaluating students' learning gains in the Alien Health game.

**Table 1: Questionnaire Dimensions, Subscale Details and Individual Items**

|   |   |  |
|---|---|--|
| <b>DIMENSION 1:<br/>Relationships</b>                 | <b>Student Negotiation (SN): a 5-item subscale assessing the extent to which students have opportunities to discuss their questions and their solutions to questions (adapted from Maor &amp; Fraser, 2005)</b> |  |
|   | SN-1  | I get the chance to talk to other students                       |
|   | SN-2  | I discuss with other students how to conduct investigations      |
|   | SN-3  | I ask other students to explain their ideas                      |
|   | SN-4  | Other students ask me to explain my ideas                        |
|   | SN-5  | Other students discuss their ideas with me                       |
|   | <b>Student Cohesiveness (SC): a 6-item subscale assessing the extent to which students are supportive to each other (adapted from Wu, Chang &amp; Guo, 2007)</b>  |  |
|   | SC-1  | Students are friendly to each other                              |
|   | SC-2  | Students are willing to help each other                          |
|   | SC-3  | It is easy to find members for grouping                          |
|   | SC-4  | Students share information with each other                       |
|   | SC-5  | Students have opportunities to discuss questions with classmates |
| SC-6  | Group members complete assignments together in class  |  |
| <b>DIMENSION 2:<br/>Personal Development</b>          | <b>Competition &amp; Efficacy (CE): a 6-item subscale assessing the extent to which students are motivated and confident to compete each other (adapted from Wu, Chang &amp; Guo, 2007)</b>                     |  |
|   | CE-1  | Students care about their own performance                        |
|   | CE-2  | Students work hard to outperform others                          |
|   | CE-3  | Classmates' performances push students to be more diligent       |
|   | CE-4  | Students set up study goals on their own                         |
|   | CE-5  | Comparisons among groups occur                                   |
|   | CE-6  | Students are confident of learning this subject well             |
|   | <b>Reflective Thinking: a 5-item subscale assessing the extent to which students have opportunities to discuss their questions and their solutions to questions (adapted from Maor &amp; Fraser, 2005)</b>      |  |
|   | RT-1  | I get to think deeply about how I learn                          |
|   | RT-2  | I get to think deeply about my own ideas                         |
|   | RT-3  | I get to think deeply about new ideas                            |
|   | RT-4  | I get to think deeply how to become a better learner             |
| RT-5  | I get to think deeply about my own understandings   |  |
| <b>DIMENSION 3:<br/>System Maintenance and Change</b> | <b>Complexity: a 5-item subscale assessing the extent to which the program is complex and represents data in a variety of ways (adapted from Maor &amp; Fraser, 2005)</b>                                       |  |
|   | C1  | It has an interesting screen design                              |
|   | C2  | It is easy to navigate   |
|   | C3  | It is fun to use   |
|   | C4  | It is easy to use  |
|   | C5  | It takes only a short time to learn how to use                   |

### Technology integration survey

Students' perceptions of the technology integration were evaluated at the end of the experience. The questionnaire was composed of five subscales guided by Moos (1987)'s conceptual framework of technology integration, later used by Wu, Chang and Guo (2007) and Maor and Fraser (2005) to derive subscales for its three dimensions: "Relationship", "Personal development" and "System maintenance and change" as in Table 1.

### Post-activity interviews

Eight students from each condition participated in an approximately 15-minute semi-structured individual interview, which took place right after the intervention. The students were asked to talk about their learning experience with Alien Health, as well as their use and perceptions of the utilized technology. Driven by Moos (1987)'s conceptual framework of technology integration and its three dimensions, the students were particularly probed to discuss the factors affecting their experience in terms of: (a) Personal development (e.g., What were the main factors that help you learn during your participation in this digital game?), (b) Relationships with others (e.g., How was the collaboration among team members structured around the digital game employed?), and (c) Technology use (e.g., Did you encounter any problems while using the digital game? How those problems affected you?). All interviews were transcribed and coded within the three dimensions of our conceptual framework.

## Findings

### Setting the baseline

A Mann-Whitney U test was used to identify any potential differences between groups in student' attitudes towards computers and digital games (Table 2). Results showed that there were no statistical differences in the student's gaming attitudes ( $U_{(40)}=198.5$ ,  $z=-.45$ ,  $p>.05$ ) and attitudes towards computers ( $U_{(40)}=183$ ,  $z=-.84$ ,  $p>.05$ ) between the groups.

Table 2: Baseline assessment of students' gaming attitudes and attitudes towards computers

|                     | Condition 1       |      | Condition 2        |      | Z     |
|---------------------|-------------------|------|--------------------|------|-------|
|                     | Kinect-based game |      | Desktop-based game |      |       |
|                     | Mean              | SD   | Mean               | SD   |       |
| Gaming attitudes    | 3.38              | 0.65 | 3.28               | 0.52 | -0.45 |
| Computers attitudes | 3.93              | 0.70 | 3.69               | 0.84 | -0.84 |

Note. \* $p\leq.05$ , \*\* $p\leq.01$ . \*\*\* $p\leq.001$

### Knowledge gains

A Mann-Whitney U test was used to identify any potential differences between groups in student' pre- and post-test scores (Table 3). Results showed that there were no statistical differences in the student's pre-test scores ( $U_{(40)}=212$ ,  $z=-.11$ ,  $p>.05$ ). However, focusing on the post-test scores, students in Condition 2, who employed the desktop-based game, outperformed their counterparts in Condition 1, who employed the kinect-based game, and this difference was statistically significant ( $U_{(40)}=139.5$ ,  $z=-1.96$ ,  $p\leq.05$ ) between the groups.

Table 3: Pre-test and Post-test scores

|                  | Condition 1       |      | Condition 2        |      | Z      |
|------------------|-------------------|------|--------------------|------|--------|
|                  | Kinect-based game |      | Desktop-based game |      |        |
|                  | Mean              | SD   | Mean               | SD   |        |
| Pre-test scores  | 1.33              | 1.49 | 1.39               | 1.69 | -0.11  |
| Post-test scores | 5.52              | 2.99 | 7.14               | 2.75 | -1.96* |

Note. \* $p\leq.05$ , \*\* $p\leq.01$ . \*\*\* $p\leq.001$

### Perceptions of technology integration

A Mann-Whitney U test was used to identify any potential differences in student's perceptions of technology integration across conditions. The results showed that students in Condition 2 had better perceptions of technology integration regarding "Student Negotiation" subscale. Yet, there were no statistical differences between the groups on all other subscales (see Table 4).

Table 4: Technology integration perceptions

|                        | Condition 1       |      | Condition 2        |      | Z      |
|------------------------|-------------------|------|--------------------|------|--------|
|                        | Kinect-based game |      | Desktop-based game |      |        |
|                        | Mean              | SD   | Mean               | SD   |        |
| Student Negotiation    | 3.36              | 1.01 | 3.97               | 0.60 | -2.20* |
| Student Cohesiveness   | 3.85              | 0.97 | 4.08               | 0.70 | -0.61  |
| Reflective Thinking    | 3.41              | 0.95 | 3.81               | 0.79 | -1.24  |
| Competition & Efficacy | 3.53              | 0.81 | 3.59               | 0.93 | -0.06  |

## Factors contributing to students' perceptions of technology integration

The thematic analysis led to the identification of numerous factors (codes) influencing students' attitudes, namely perceptions of technology integration, within the three dimensions of our conceptual framework: "Relationship", "Personal development" and "System maintenance and change". A conceptual map was the result of further organizing the emerging factors into basic themes: (a) Content-related factors, referring to the features of the gaming content, (b) Interface-related factors, referring to the affordances of the gaming platform, (c) Activity-related factors related to the pedagogical setting in which the game was contextualized and (d) Context related factors, referring to the characteristics of the physical environment in which the activity was enacted. All factors were evaluated as positive or negative in relation to their impact on students' perceptions (see Table 5).

Table 5: Categorization of factors reported as affecting students' perceptions on technology integration

| Framework Dimensions          | Basic themes              | High-embodied condition<br>[Kinect-based game] |                            | Lowly-embodied condition<br>[Desktop-based game] |                        |
|-------------------------------|---------------------------|--|----------------------------|--|------------------------|
|                               |                           | Positive factors (+)                           | Negative factors (-)       | Positive factors (+)                             | Negative factors (-)   |
| Personal development          | Content related factors   | Learning content                               | Textual information        | Learning content                                 | Textual information    |
|                               |                           | Gaming features                                |                            | Gaming features                                  |                        |
|                               |                           | Narrative plot                                 |                            | Narrative plot                                   |                        |
|                               |                           | Integrated scaffolding                         |                            | Integrated scaffolding                           |                        |
|                               | Activity related factors  | Worksheets                                     | Gaming nature              | Worksheets                                       | Gaming nature          |
|                               | Interface related factors | Embodied interactions                          | Locomotion                 |  |                        |
| Context related factors       |                           | Classroom noise                                |                            | Classroom noise                                  |                        |
|                               |                           | Other group interventions                      |                            | Other group interventions                        |                        |
| Relationship                  | Activity related factors  | Team-based mode                                | Large groups               | Team-based mode                                  |                        |
|                               |                           | Collaborative writing task                     | Unstructured collaboration | Collaborative writing task                       |                        |
|                               |                           | Peer feedback strategies                       | Waiting time               | Peer feedback strategies                         |                        |
|                               | Interface related factors |  | Single-player mode         |  | Single-player mode     |
| System Maintenance and Change | Interface-related factors | Novel interface                                | Gaming controls            | Gaming controls                                  | Small projection       |
|                               |                           | Large projection                               | Synchronization issues     |  | Low graphics interface |
|                               |                           | Bodily movement                                |                            |  |                        |
|                               |                           | Touchless interaction                          |                            |  |                        |

### Personal development

Students in both conditions reported how a set of content-related factors such as the *learning nature* of the game, the *gaming features* (e.g. stages, points, rewards), the *narrative plot* on which the game was structured, as well as the *integrated scaffolding* (e.g., hints and prompts) had a positive contribution to their personal learning development. E.g.

*"I liked the game's narrative plot as there was an alien trying to go back to his planet. We had to feed the alien with healthy foods. I liked the fact that every new planet was a new stage in the game with a new activity to do. It was an educational game because you could learn about nutrients in food."* [#Girl -L-, Desktop-based version]

However, students in the high-embodied cognition highlighted that there was too much *textual information*, while students who worked in the lowly-embodied condition also added that there was a *repetition of the gaming stages*, which in turn had a negative impact of their interest. E.g.

*“A negative factor in the game that I can think of was the large text. A box popped up in every new stage writing a lot of text. Text could be limited.”* [#Girl -E-, Kinect-based version]

Focusing on the activity-related factors, students in both conditions, added that while the *worksheets* that they were required to complete contributed positively to their personal development, the *gaming nature* sometimes inhibited the learning process, as in many cases the students would deal with the activity as a playful rather than as an educational experience. E.g.

*“We were carried away when playing the game and were oftentimes forgetting to complete our paper assignment. It was difficult to remember later on what to write.”* [#Girl -L-, Desktop-based version]

Focusing on the interface-related factors, students in the high-embodied condition reported that while the activity allowed for *embodied interactions* which were valuable for their learning and personal development, *locomotion* was in some cases a negative aspect. In particular, as some of students admitted, in some cases they would be more focused on coordinating their body movements, rather than on the learning content. E.g.

*“Moving my body did not help me being concentrated but I had a lot of fun. If I was stable in front of a computer I would have been more concentrate because I would focus on the screen and click the correct answer, rather than trying to coordinate my body.”* [#Girl -E-, Kinect-based version]

Finally, students in both conditions reported how the *classroom noise* and *other groups’ interventions* while working, were two main context-related factors negatively affecting their personal development.

### Relationship

Students’ in both conditions reported how a set of activity-related factors such as the *team-based mode* in which the activity was enacted, the *collaborative writing task* that were assigned (one worksheet to be completed by each group) and the *peer feedback strategies* that were followed, had a positive impact on their collaboration. More specifically, as the students mentioned all these factors promoted productive social interactions, such as exchange of views and ideas, peer scaffolding and assistance. E.g.

*“I liked working in my group to complete the paper assignment. We worked collaboratively and we were helping each other. We were helping our co-players to choose the correct answer, we were giving instructions and we were encouraging each other to try harder.”* [#Girl -E-, Kinect-based version]

However, students in the high-embodied condition also negatively elaborated on how a set of activity-related factors, such as working in *large groups* (of 4 students) in combination to the *unstructured collaboration*, affected their relationships negatively. In particular, as the students admitted, both of these factors prohibited their effective collaboration, as it was more difficult to agree on a common strategy and plan their next steps, while there were also many disagreements with children often fighting over turn-taking and roles in the group. E.g.

*“I wanted to play more but the other members in my team urged me to finish so they could play. There was also a boy taking my turn in the game. He wanted to play instead of me. I couldn’t concentrate because my team members were telling me the correct answers, or they were trying to show me how to move. I got confused!”* [#Boy -M-, Kinect-based version]

Finally, students in both conditions, highlighted that in terms of the interface-related factors, the *single-player mode* of the game, transformed the non-player(s) as spectators, and this had also a negative effect on students’ relationships. Importantly in the high-embodied condition, this factor had an increased negative effect given the increased *waiting time* between turns, which resulted in off-task discussions and behaviors amongst the members of the group. E.g.

*“The game was for a single player. All the other members of the group stayed aside, they had conversations with each other about topics unrelated to the game’s content and they were not concentrated in their team members’ actions neither on contributing to the group’s collaboration.”* [#Girl -A-, Kinect-based version]

### System maintenance and change

According to the students of the high-embodied condition, the *large projection* (bigger screen providing more heightened sensory stimuli), the *interface* (with the use of novel technologies), as well as the affordances of the

gaming platform for promoting *bodily movement* (via the gesture-based interactions), contributed to their experienced immersion and this had a positive impact on their perceptions of technology use. E.g.

*“There was a large screen which seemed nicer and easier. I could see everything in that big screen. I could have better control of the game and I could feel like being in the game!”* [#Boy -I-, Kinect-based version]

However, the students of the high-embodied condition reported that the *controls* of the game which were rather different from traditional gaming controls, some *synchronization* issues often presented between students’ movements and their belated projection on the screen, as well as some technical bugs (provoked by students’ proximity to the Kinect), affected their perceptions of technology in a negative way. E.g.

*“Sometimes there were problems with the technology. The game blocked and our hand signal was not appearing on the screen or was presented in a wrong position. This cost us time as we had to wait for the problem to be resolved!”* [#Boy -I-, Kinect-based version]

Finally, students in the low-embodied cognition reported that the small projection (limited desktop screen) and the low graphics *interface* had a negative effect on their perceptions about the technology use. However, students in the low-embodied condition explained that the game had familiar *gaming controls* (keyboard and mouse) and thus, was more easily integrated in the lesson.

## Discussion and implications

The present investigation examined students’ learning and perceptions of technology integration for group learning in a high-embodied, Kinect-based educational game, in comparison with a low-embodied, desktop-based version of the same game, in an authentic classroom setting. Findings suggest higher learning gains and more positive perceptions of technology integration, in terms of opportunities for student negotiation, for the students in the low-embodied condition.

*Are there differences in student’s leaning outcomes and attitudes between the conditions?* In the present work, there was no difference in most dimensions of students’ attitudes, namely, perception of technology integration, across conditions. Yet, the opportunity for “student negotiation” was deemed higher in the low-embodied condition. Also, students in the low-embodied condition presented increased knowledge gains in comparison to their counterparts in the high-embodied condition. In general, the results contradict findings of prior research conducted in laboratory settings in which the prevalence of high-embodied, over low-embodied games, in students learning is presented (Homer et al., 2014; Lindgren et al., 2016). Indeed, the present study supports previous evidence (from a limited number of studies conducted in authentic school classrooms) that the high-embodied experience has not been as successful as initially expected in promoting students’ learning compared to low-embodied, desktop-based environments (e.g., Anderson & Wall, 2016; Hung, Lin, Fang & Chen, 2014). That is, while being enjoyable and engaging, the experience with embodied learning technologies used in a typical classroom environment to run learning tasks may not always produce significant learning gains.

*What are the main factors affecting students’ perceptions of technology integration in the two conditions?* The analysis of students’ post-activity interviews shed light to our findings; a series of contextual factors were mentioned by students to have affected their perceived experience in a negative way. For example, common technical issues or a noisy classroom environment, may detract from, rather than enhance student learning, which is not a surprising result (e.g., Darling-Aduana & Heinrich, 2018). The study presented a conceptual map to summarize these factors into content-related factors, interface-related factors, activity-related factors and context related factors affecting the experience, positively or negatively, in both conditions. The map can be informative in future research and practice in the area allowing to control for some of these factors in the authentic learning environment.

Evaluating the outcomes of the present case study, several limitations should be noted. Conducting scientific research in a functioning classroom environment was challenging which naturally introduced flaws in the implementation of the study. For example, the overall time for the activity was fixed by the school time-table which did not allow much time for familiarization with the game mechanisms, especially in the high-embodied (Kinect-based) condition. The classroom’s setting imposed several additional constraints; having a number of students in groups interacting with Kinect cameras created undesirable noise and interference. There is clearly much more work that could be done to explore the best ways of integrating embodied learning technologies within a classroom setting. Given the popularity of embodied leaning technologies in the recent days, this work helps to identify issues which warrant future investigation in the field of emerging technology integration. A central question remains to be answered: under what circumstances can embodied technologies be educationally beneficial in authentic classroom settings? Future research may wish to focus on specific strategies for embodied

learning technology integration, designed to be immediately adopted by in-service teachers. Such work could be beneficial in increasing integration of emerging technology promising to enable students' learning gains and positive perceptions of embodied technology use for learning in groups.

## References

- Abrahamson, D., & Sánchez-García, R. (2016). Learning is moving in new ways: The ecological dynamics of mathematics education. *Journal of the Learning Sciences*, 25(2), 203-239.
- Anderson, J. L., & Wall, S. D. (2016). Kinecting physics: Conceptualization of motion through visualization and embodiment. *Journal of Science Education and Technology*, 25(2), 161-173.
- Bonanno, P., & Kommers, P. A. (2008). Exploring the influence of gender and gaming competence on attitudes towards using instructional games. *British Journal of Educational Technology*, 39(1), 97-109.
- Bressler, D. M., & Bodzin, A. M. (2013). A mixed methods assessment of students' flow experiences during a mobile augmented reality science game. *Journal of Computer Assisted Learning*, 29(6), 505-517.
- Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research*. Sage publications.
- Darling-Aduana, J., & Heinrich, C. J. (2018). The role of teacher capacity and instructional practice in the integration of educational technology for emergent bilingual students. *Computers & Ed*, 126, 417-432.
- Georgiou, Y., Ioannou, A.: Embodied learning in a digital world: A systematic review of empirical research in K-12 education. In: Díaz, P., Ioannou, A., Spector, M., & Bhagat K.-K.(eds.) Learning in a digital world: A multidisciplinary perspective on interactive technologies for formal and informal education. Springer series: Smart Computing and Intelligence. Springer (In press).
- Homer, B. D., Kinzer, C. K., Plass, J. L., Letourneau, S. M., Hoffman, D., Bromley, M., ... & Kornak, Y. (2014). Moved to learn: The effects of interactivity in a Kinect-based literacy game for beginning readers. *Computers & Education*, 74, 37-49.
- Hung, I. C., Lin, L. I., Fang, W. C., & Chen, N. S. (2014). Learning with the body: An embodiment-based learning strategy enhances performance of comprehending fundamental optics. *Interacting with Computers*, 26(4), 360-371.
- Johnson-Glenberg, M. C., & Hekler, E. B. (2013). "Alien Health Game": An embodied exergame to instruct in nutrition and MyPlate. *Games for Health: Research, Dev, and Clinical Applications*, 2, 354-361.
- Karakostas, A., Palaigeorgiou, G., & Kompatsiaris, Y. (2017, November). WeMake: A framework for letting students create tangible, embedded and embodied environments for their own STEAM learning. In *International Conference on Internet Science* (pp. 3-18). Springer, Cham.
- Kosmas, P., Ioannou, A., & Retalis, S. (2018). Moving Bodies to Moving Minds: A Study of the Use of Motion-Based Games in Special Education. *TechTrends*, 1-8.
- Kosmas P., Ioannou A., Retalis S. (2017). Using Embodied Learning Technology to Advance Motor Performance of Children with Special Educational Needs and Motor Impairments. In: Lavoué É., Drachler H., Verbert K., Broisin J., Pérez-Sanagustín M. (eds) Data Driven Approaches. EC-TEL 2017. Lecture Notes in CS, vol 10474. Springer.
- Lindgren, R., Tscholl, M., Wang, S., & Johnson, E. (2016). Enhancing learning and engagement through embodied interaction within a mixed reality simulation. *Computers & Education*, 95, 174-187.
- Malinverni, L., & Pares, N. (2014). Learning of abstract concepts through full-body interaction: A systematic review. *Educational Technology & Society*, 17 (4), 100–116.
- Maor, D., & Fraser, B. J. (2005). An online questionnaire for evaluating students' and teachers' perceptions of constructivist multimedia learning environments. *Research in Science Education*, 35(2-3), 221-244.
- Melcer, E. F., & Isbister, K. (2016, May). Bridging the physical divide: A design framework for embodied learning games and simulations. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 2225-2233). ACM.
- Moos, R.H. (1987). *The social climate scales: A user's guide*. Palo Alto, California: Psychologists Press.
- Teo, T., & Noyes, J. (2008). Development and validation of a computer attitude measure for young students (CAMYS). *Computers in Human Behavior*, 24(6), 2659-2667.
- Wilson, M. (2002). Six views of embodied cognition. *Psychonomic bulletin & review*, 9(4), 625-636.
- Wu, W., Chang, H. P., & Guo, C. J. (2009). The development of an instrument for a technology-integrated science learning environment. *International Journal of Science and Mathematics Education*, 7(1), 207.

## Acknowledgments

This work is part of a project that has received funding from the European's Horizon 2020 research and innovation programme under grant agreement N°739578 and the government of the Republic of Cyprus through the Directorate General for European Programmes, Coordination and Development.

# Effective Regulation in Collaborative Learning: An Attempt to Determine the Fit of Regulation Challenges and Strategies

Nadine Melzner, Martin Greisel, Markus Dresel, and Ingo Kollar  
nadine.melzner@phil.uni-augsburg.de, martin.greisel@phil.uni-augsburg.de, markus.dresel@phil.uni-  
augsburg.de, ingo.kollar@phil.uni-augsburg.de  
University of Augsburg

**Abstract:** University students often self-organize in groups for collaborative exam preparation. To learn effectively, they need to regulate emerging learning challenges through the choice of fitting strategies. So far, little is known about what strategies fit which challenges. We present a theoretical account for the definition of fit between regulation challenges and strategies in collaborative learning, thereby differentiating between direct and indirect strategies, and test its validity in an empirical study. We asked 163 university students in 90 groups to rate the challenges they encountered in their self-organized group meetings, and to report strategies by aid of which they regulated their biggest challenge. Answers were coded into 26 strategy types. We found that students mostly used strategy types that directly address their biggest challenges. Multilevel-modelling indicated that direct strategies are associated with higher satisfaction for coordination- and comprehension-related challenges only, but not for motivational challenges.

**Keywords:** regulation challenges, collaborative learning, regulation fit, exam preparation, self-regulated learning

## Problem statement

Many university students choose to prepare for exams in groups. Potential benefits of this are to maintain one's own discipline or to improve one's own understanding by discussing learning material (Järvenoja, Volet, & Järvelä, 2013). Yet, groups often face coordination-related, motivational, or comprehension-related challenges. When such challenges are not regulated, group members may be dissatisfied and reach poor group outcomes (Freeman, 1996). However, little is known about what strategies actually fit what kinds of challenges that may occur during collaborative learning and how their application relates to learners' satisfaction with their learning process. We (a) provide a theory-based analysis of the fit between learning challenges and regulation strategies and (b) investigate whether this fit is predictive for learners' satisfaction with their group learning experience.

## Theoretical Background

### Types of challenges

We propose that most challenges that may occur during collaborative learning fall into one of the following three categories: (a) coordination between group members, (b) motivational aspects, or (c) comprehension of learning topics and materials. *Coordination-related challenges* may arise when group members set different priorities in learning the exam contents, or when they set different learning goals (Järvelä & Järvenoja, 2011). *Motivational challenges* may emerge when group members, for example, cannot find interest in the exam topics. Finally, *comprehension-related challenges* may arise when groups have too little subject-specific prior knowledge, or when the learning material is structured in a confusing way. For successful exam preparation, learners need to effectively regulate these challenges. So far, studies (e.g., Malmberg, Järvelä, Järvenoja, & Panadero, 2015) have tried to predict the regulatory success by the quantity or quality of strategy use. Yet, in line with theoretical models of self-regulated learning (e.g., Zimmermann & Moylan, 2009), we additionally suppose that in order to be effective, there is the need for a chosen strategy to match the specific challenge at hand.

### Strategy fit

The fit of strategies to challenges has been described as an important factor that influences the success of learning regulation (Engelschalk, et al., 2015). Nevertheless, in collaborative learning, research about which strategy types are effective in what situation, i.e., what strategy types "fit" a given challenge, is scarce. Therefore, in this paper, we make an attempt to determine the appropriateness of specific strategy types to regulate specific coordination-related, motivational, and comprehension-related challenges.

We firstly assume that in a given situation some strategies may fit a current regulation challenge better than others. In one situation, group learners might for example differ in their ideas on how to proceed with the task (= a specific coordination-related challenge): One group member might want to recapitulate a topic that was already covered by the group, while other group members might prefer to improve their comprehension of a new topic. Now the group needs to react to the challenge by choosing a regulation strategy. It could, for example, apply a motivation strategy like planning to reward themselves after the learning session, or it could discuss these divergent plans and find an agreement on how to proceed. For the given situation, the latter strategy seems to be more effective to overcome the given challenge than the aforementioned strategy because it aims at directly eliminating the challenge. At least, communicating own plans and intentions is necessary to negotiate the further and joint proceeding (= to resolve the challenge). This does not mean at all that the reward strategy necessarily is inappropriate in this learning situation as a whole. Actually, it might support learning at a more general level because it might address different (motivational) problems that might be present in the same learning situation simultaneously. Nevertheless, it would rather not or not effectively help to dissolve the specific challenge of divergent plans. Hence, strategies that can help to *directly* eliminate a given challenge might be more likely to result in effective regulation than strategies that might support the learning process in a different, but non-direct way that leaves the central challenge itself unaffected. Thus, we regard the choice of strategies that directly aim at eliminating the challenge as “fitting”.

Other challenges, e.g., “distraction”, can be overcome by multiple strategy types. In this case, specific resource-oriented strategies (e.g., controlling the environment to reduce external stimuli), metacognitive strategies (discussing how to organize learning to hinder distraction), or a selection of motivational strategies (e.g., making the topic more fun to learn, i.e. increasing situational interest; see Table 1) seem promising, as they all address the problem in a direct manner (e.g., when any of these strategies is applied successfully, the problem should disappear). We assume that the amount of strategy types considered as being effective varies with the specificity or globality of the perceived challenge: Coordination-related and comprehension-related challenges seem more specific, thus less strategy types directly address them. Motivational challenges are more global and thus can be addressed by more different strategy types. Table 1 provides an overview about which strategy types theoretically can be considered to directly regulate different kinds of regulation challenges based on the considerations above.

## Relation of adaptive strategy use and satisfaction with the learning experience

One (among several) way(s) to determine the effectiveness of specific regulation strategies is through measuring learners’ satisfaction with the overall learning experience. This assumption can be derived from the cyclic phase model of self-regulated learning (Zimmerman & Moylan, 2009): In terms of this model, satisfaction is a self-reaction which depends on the goals that were set in the beginning of a learning activity (or the evaluation standards in the model of Winne, Jamieson-Noel, & Muis, 2001). Concerning self-initiated study groups in preparation for an exam, one may assume that the main goal these students set for their meetings is to learn effectively (and eventually to receive good grades). Thus, the (subjectively) acquired knowledge that is accumulated during collaborative learning may be understood as a measure for learning success in the self-reflection phase. If the goals/standards are met, satisfaction should result.

Even though the assumption that regulation strategies need to fit the regulation challenges to yield effective learning is immanent in models of self-regulated learning, it has rarely been empirically tested. In one study, Järvelä, Volet, and Järvenoja (2010) filmed the regulatory activity of 63 teacher education students during collaborative learning tasks, and had participants rate the perceived intensity of various emotional challenges they experienced during group work as well as their satisfaction with their group learning process. Using qualitative analyses, they showed that one group demonstrated effective social and self-regulation of different challenges, and, at the same time, rated its group learning as “relatively satisfactory” up to “very satisfactory”. Even though this study investigated regulation fit only for emotional challenges, it suggests a link between the problem-specific fit of regulatory strategies and learners’ satisfaction with the learning experience. If we find that learners who apply strategies that based on our theoretical analysis directly address their challenges are more satisfied with their learning than those learners who apply other strategies, this result could be regarded as evidence that is in line with our theoretical assignment of strategies to challenges.

## Research questions and hypothesis

We ask the following research questions: To what extent do group learners use strategies that directly address a reported challenge compared to other strategies (RQ 1)? Can this extent be considered to be sensitive to the challenge type, i.e., is it more likely that direct compared to other strategies are chosen (RQ 2)? Are learners who apply direct strategies more satisfied with their learning than those who used other strategies and do differential effects depend on the problem type being regulated (RQ 3)? RQ 1 and RQ 2 will be answered in an exploratory



**Table 1: Specific strategy types for direct regulation of specific coordination-related, motivational, and comprehension-related challenges**

| Specific challenge  | Strategies directly addressing the challenge |                     |   |                            | Strategy example  |
|---|--|---------------------|---|----------------------------|---|
|   | Cog-<br>nitive                               | Metacog-<br>-nitive | Motiva-<br>tional                               | Re-<br>source-<br>oriented |   |
| <b>Coordination-related Challenges</b>  |  |                     |   |                            |   |
| A. The group members have different goals for the meeting.                            |  | PRL                 |   | EM                         | PRL: "I advised her to take a closer look at it at home, as I do not want it that much in detail."<br>PRL: "I gave them tips for learning." |
| B. The group members seem to have incompatible working styles.                        |  | PRL                 |   |                            | (No examples available)   |
| C. The group members seem to have different communication styles.                     |  | PRL                 |   |                            | (No examples available)   |
| D. The group members understand concepts / tasks differently.                         | RDU  | PRL                 |   |                            | RDU: "I have justified my conception of the task. "   |
| E. The group members have different ideas on how to proceed with the task.            |  | PRL                 |   |                            | PRL: "We discussed how we want to proceed."   |
| F. The contributions of individual group members are differently strongly considered. |  |                     |   | EM                         | (No examples available)   |
| G. Not every group member always dares to participate.                                |  |                     | ASI   | EM                         | EM: "I participated in the study group."  |
| I. Not all group members have enough time for the meeting.                            |  |                     |   | TMC, KIM, AM               | TMC: "I tried to use the time as effectively as possible through targeted questions."   |
| <b>Motivational Challenges</b>  |  |                     |   |                            |   |
| H. The group has distraction problems.  |  | PRL                 | SIT, DSG, ACC, FC                               | EC, AM                     | EC: "I have asked my classmates to turn off all phones."  |
| K. The group members have motivational problems.                                      |  |                     | RS, SIT, IPS, LPS, ASI, DSG, AFC, HUG, ECG, UMS |                            | RS: "I told her that there is only little content left and that we have already done most of the work"                                      |
| M. The group members consider the study material to be boring.                        |  |                     | RS, SIT, IPS, LPS, DSG, AFC, HUG, ECG           |                            | ESG: "Tried to keep them at it."  |
| <b>Comprehension Challenges</b>   |  |                     |   |                            |   |
| J. The group members have only low prior knowledge of the learning contents.          | SIC, CGP, ERM                                |                     |   |                            | SIC: "The mutual explanations were helpful for the whole group"   |
| L. The group perceives the learning material as difficult.                            | SIC  |                     |   | TMC                        | SIC: "Tried to together clarify unclear contents"   |
| N. The group members perceive the study material as confusing.                        | ORR, SIC                                     |                     |   |                            | ORR: "We have created a mind map to structure the topic."   |

*Note. Cognitive:* Organizational / Knowledge Reduction Strategies / Knowledge Recall Strategies (ORR), Strategies for Improving Comprehension (SIC), Strategies for Closing Gaps in Prior Knowledge (CGP), Strategies to Resolve Differences in Understanding (RDU), Surface-oriented strategies / Knowledge Retention Strategies / Consolidation Strategies (SRC)

*Metacognitive:* Planning and Regulation of the Learning Process (PRL), Reflection and Evaluation of the current state of Knowledge (REK)

*Motivational:* Reward Strategies (RS), Increasing Situational Interest (SIT), Increasing Personal Significance (IPS), Learning and Performance-related Self-instruction (approach and avoidance) (LPS), Ability-related Self-Instruction (ASI), Declaring Successful Self-control as Goal (DSG), Accentuating Frame Conditions or Constraints (AFC), Highlighting Group Utility as a Goal (HUG), Emotional Contagion (ECG), Unspecific Motivational Strategies (UMS)

*Resource-oriented:* Time Management and Coordination (TMC), Environmental Control (EC), Knowledge and Information Management (KIM), Attention Management (AM), Effort Management (EM), External Resource Management (ERM), Care of the Social Atmosphere (CSA).

fashion. With regard to RQ 3, we hypothesize that learners who apply direct strategies to remedy their biggest regulation challenge are more satisfied with their group learning than students who use other strategies. Yet, we additionally explore whether there are differential effects of the problem type, i.e. that direct strategies are only related to satisfaction when addressing certain, but not other kinds of challenges.

## Method

### Sample

Participants were  $N = 174$  university students (74.29% female, 25.14% male, 0.57% other) belonging to 90 self-organized study groups. They had an average age of 22 years ( $M = 22.26$ ,  $SD = 2.68$ ) and were, on average, in their 5<sup>th</sup> semester ( $M = 4.64$ ,  $SD = 3.03$ ) of their current studies and in their 6<sup>th</sup> ( $M = 6.21$ ,  $SD = 3.85$ ) semester of studies in general. Participants were enrolled in various study programs (e.g. teacher education, law, mathematics, computer science) at different German universities. The number of students per group participating in the study ranged from 1 to 5, with two participating group members being the most frequent fraction (23.27%). After each group meeting, participants were asked to (individually) answer an online questionnaire. Participants were reimbursed with 14 € if they filled in the questionnaire 3 times. For each additionally completed questionnaire, they received an additional 2.50 €. The maximum was 10 questionnaires (35 €).

### Instruments

#### Questionnaire

The used questionnaire was a modified version of the Adaptive Instrument for Regulation of Emotions (“AIRE”; Järvenoja & Järvelä, 2009). It asked for the experience and regulation of coordination-related, motivational, and comprehension-related challenges during the current group meeting. Participants first were asked to rate on a five-point scale (1= “does not apply at all”, 5= “does squarely apply”) how intensively they experienced each of the 14 listed challenges during their group meeting (see Table 1). Then, they should mark their biggest of those 14 challenges and describe it briefly. This was followed by an item that asked students to rate their satisfaction with the respective group meeting on a seven-point rating scale (1= “totally unsatisfied”, 7= “totally satisfied”). In the end, participants were asked to enter strategies they used to regulate their biggest self-categorized challenge at the self-, co-, and shared level in an open answer format (e.g. self: “What did you personally think, do or say to ensure high quality of your own learning in this situation?”, co: “...of the learning of individual others”, shared: “...of the learning of the group as a whole”; Järvelä & Hadwin, 2013).

#### Coding procedure

After data collection, a coding scheme to classify learning strategies was developed, based on established schemes of Mandl and Friedrich (2006), and of Engelschalk et al. (2015). It contained 26 categories: Twenty-four specific strategy types (see “Note” in Table 1 for a complete list), one residual category for unspecific strategies, and one category when no strategy was provided. Trained coders’ categorization yielded a sufficient interrater agreement (Cohen's  $\kappa = .85$ ). After coding, a binary “fit” variable was generated by assigning the values 1 or 0 based on the aforementioned theoretical scheme (1 if a person mentioned at least one strategy directly addressing the perceived biggest challenge, 0 if only other strategies were mentioned; see Table 1 for details about which strategies theoretically match which challenges).

### Analysis

To answer RQ 1, we calculated absolute and relative frequencies of the use of the different strategies (together with 95% confidence intervals) and visualized them with a decision tree (see Figure 1). Concerning RQ 2, we added ratios “R” for each challenge type: To give an impression of how to interpret the magnitude of the relative frequencies of strategies directly addressing a challenge to other strategies (= whether regulation is adapted to a specific challenge), we calculated ratios of the empirical to the theoretical relations of direct and total strategy use. The theoretical relation signifies the probability to choose a direct strategy, if a random distribution across all strategy categories was assumed, weighted with the average number of strategies reported for each challenge type. In summary, this ratio expresses the empirical probability that a person directly addresses a specific challenge qualified by what would be expected if there was no adaption in the choice of regulation strategy to a challenge at all. Values above 1 signify that more participants actually chose direct strategies than one would expect under the assumption of random distribution of strategies, whereas values below 1 signify that less participants chose direct strategies than one would expect if there was no adaption to challenge. The deviation from 1 was tested by asymptotical  $\chi^2$ -tests.

To test if choosing a strategy type that fits a current challenge increases the satisfaction with learning (Hypothesis 1), hierarchical linear modelling using R (reml estimation) was conducted. Data was analysed using a three-level data structure (group meetings, persons, groups; Singer & Willett, 2003). Students' satisfaction with their learning and group differences were estimated by specifying the *unconditional means model* where satisfaction with learning is independent of predictors. Nevertheless, we allowed for random variations in satisfaction between groups and individuals in groups (random parameters  $\tau_0$ ,  $u_{00}$ ). We z-standardized the satisfaction variable to interpret differences in satisfaction in units of standard deviation. General development of students' satisfaction with their learning across group meetings and group differences in these developments were estimated by specifying the *unconditional growth model*. In this model, satisfaction with learning is modeled as dependent on the number of learning sessions ( $\gamma_{001}$ ). This time effect is allowed to vary randomly between individuals and between groups (random parameters  $\tau_0$ ,  $u_{00}$ ). The group meeting was coded by number to approximately interpret the parameters of the time variable as changes per meeting (approximately because some groups already existed before our study). Then, we specified a *fit model* that included the effects mentioned above and additionally the effect of the fit between the challenge and strategy type on satisfaction. The fit variable was included as a fixed effect because there is no plausible argument for random variation over groups and individuals in groups. Last, we added the *interaction of fit and problem type* to the last model.

## Results

### Preliminary analysis

We first looked at whether the types of challenges we presented to participants actually fell into the three assumed categories "coordination-related", "motivational", and "comprehension-related" challenges. To judge this, we used the individual ratings to what extent the 14 challenges actually occurred in the group in exploratory and confirmatory factor analyses. The EFA supported the choice of 3 factors as the optimal factor structure by 3 (out of 9, 33.33%) methods (Optimal Coordinates, Parallel Analysis, VSS Complexity 2). A further CFA indicated a superior model fit for the three-factor structure compared to theoretically plausible one- or two-factor structures. However, the overall model fit was not optimal ( $\chi^2=2965.899$ ,  $p < .001$ ,  $\chi^2/df = 32.59$ , CFI=0.89, RMSEA=0.07, SRMR=0.06). Since the following calculations are performed separately for each specific challenge, and grouping into larger problem dimensions was used for clarity of visualization only, the decision for a three-factor structure did not pose a problem for later analyses.

### Distribution of fit between learning challenges and regulation strategies

Concerning RQ 1, Figure 1 shows the transition probabilities with which participants who encountered a specific challenge type chose a direct strategy. It also provides information on whether a larger or smaller proportion of participants regulated a specific challenge with a direct versus with other strategies. It can be seen that direct strategies were more frequent than alternative strategies across all types of challenges except for "boring learning material" for which the confidence interval includes equal frequency (50%). Overall, these results suggest that students predominantly choose strategy types for directly controlling specific challenge types, indicating high fit between selected strategies and regulation challenges. The exception for "boring learning material" was surprising since our theoretical model assumed a whole set of different strategy types to directly address this challenge type. Although the baseline probability of choosing a direct strategy was rather high, students chose other types of strategies equally often here. In figure 1, the ratios "R" signifying the relation of the empirical percentage of direct strategies to what could be expected under the assumption of a random distribution of strategies across all coding categories demonstrate values significantly (if testable due to low cell frequencies) greater than 1 for eleven out of 14 challenge types (coordination problems were not listed individually in the tree for lack of space). This hints towards a rather high ability of students to adaptively select direct strategies to remedy their regulation challenges (RQ 2). Though, the ratios of the three motivational challenges were significantly below 1, indicating regulation to be sensitive to the kind of challenge as well, but in an unexpected direction. For these problem types, students chose direct strategies less frequently and other strategies more frequently than expected when strategy selection would be random. In order to test whether the observed regulation differences between coordination- and comprehension-related problems on the one hand and motivational problems on the other hand also show up in regard to satisfaction, this split of problems was maintained in the multi-level models. So, we suppose that direct strategies are positively associated with satisfaction only for coordination- and comprehension-related challenges, whereas for motivational challenges, they might be negatively associated.

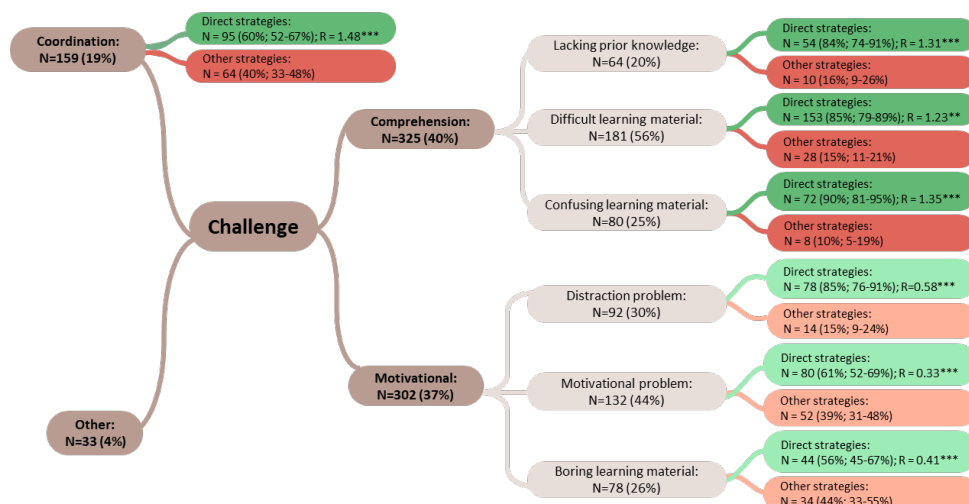


Figure 1. Decision tree with (relative) frequencies of challenges and direct vs. other strategy types (with 95% confidence intervals). Ratios “R” show the empirical probability for directly addressing a specific challenge compared to the probability expected with no adaptation in the choice of regulation strategy to a challenge. \*  $p \leq .05$ , \*\*  $p \leq .01$ , \*\*\*  $p \leq .001$ .

### Relation of fit with outcome

Regarding RQ 3, the *unconditional means model* (see Table 2) showed that 19% of the variance of satisfaction was explained by the group level. Thus, groups differed in the effectiveness of regulation. The *unconditional growth model* retained this effect but showed a main effect of time ( $\gamma_{100}$ ) indicating that satisfaction decreased across learning sessions in average. Further, the *fit model* showed the effect of time again ( $\gamma_{100}$ ) but yielded no significant effect for fit ( $\gamma_{200}$ ). Thus, the fit between the biggest challenge and the selected strategy type to regulate this challenge did not significantly predict satisfaction. The *interaction of fit and problem type model* revealed a main effect for problem type ( $\gamma_{300}$ ) indicating that group learners were more dissatisfied with their learning when facing motivational instead of coordination- or comprehension-related problems. This model further revealed a significant (one-sided according to hypothesis stated above) interaction effect between fit and problem type ( $\gamma_{400}$ ). Thus, for coordination- and comprehension-related challenges, satisfaction was higher when direct instead of other strategies were chosen, but was lower for motivational challenges.

Table 2: Outcomes of multi-level models

|                                 | Satisfaction with the learning |                            |                  |   |
|---------------------------------|--------------------------------|----------------------------|------------------|---|
|                                 | Unconditional means model      | Unconditional Growth Model | Fit Model        | Fit, Problem type and Interaction Model |
| Fixed Effects                   |                                |                            |                  |   |
| Intercept $\gamma_{000}$        | -0.030                         | 0.116 (0.078)              | 0.115 (0.078)    | 0.107 (0.078)                           |
| Time $\gamma_{100}$             |                                | -0.037** (0.011)           | -0.036** (0.011) | -0.036** (0.011)                        |
| Fit $\gamma_{200}$              |                                |                            | 0.044 (0.034)    | 0.042 (0.034)                           |
| Problem type $\gamma_{300}$     |                                |                            |                  | -0.083* (0.035)                         |
| Problem type*Fit $\gamma_{400}$ |                                |                            |                  | -0.061* (0.033)                         |
| Random Parameters               |                                |                            |                  |   |
| Level 2 (Persons)               |                                |                            |                  |   |
| Intercept Var( $\Gamma_0$ )     | 0.053*                         | 0.048*                     | 0.050*           | 0.049*                                  |
| Level 3 (Groups)                |                                |                            |                  |   |
| Intercept Var( $\nu_{00}$ )     | 0.193***                       | 0.205***                   | 0.201***         | 0.197***                                |

Note. Standard errors are in parentheses. Time = the number of study group meeting (Level 1), Fit = the fit between biggest challenge and the chosen strategy type to regulate this challenge (Level 1). \*\*\*  $p \leq 0.001$ , \*\*  $p \leq 0.01$ , \*  $p \leq .05$ .

## Discussion

The aims of our study were twofold. First, we wanted to investigate to what extent group learners who encounter specific coordination-related, motivational, and comprehension-related challenges during their exam preparation select strategy types that directly address those challenges. In this context, we also reflected on whether this extent can be considered to be sensitive to the challenge type. Second, we tried to validate our postulated fit categorization by regressing differences in satisfaction with the learning experience on the fit of the strategy types students selected with the biggest learning challenge they experienced during learning.

As expected, learners in groups particularly seem to choose strategy types directly addressing their biggest challenge. For thirteen out of 14 challenge types, they chose significantly more direct than other strategies. When additionally controlling for baseline probability of the measurement instrument, they still regulated eleven out of 14 challenges sensitive to the specific challenge. Thus, group learners seem to be able to adaptively regulate their learning challenges dependent on the kind of challenge they are confronted with. In line with the SRL model of Zimmerman and Moylan (2009), it may be that learners choose strategies that match their experienced challenges because they have turned out to be most effective for their control. However, this result seems to contradict observations of Engelschalk et al. (2015) who found that students did not differentiate much in their strategy choices and rather activate proven strategies repeatedly across situations. The diverging findings might be related due to the different study designs or contexts: In the study by Engelschalk and colleagues (2015), individual learning students were presented with hypothetical motivational problems and asked to describe how they would regulate each of them. Contrary to that, we asked group learners to report their real experienced coordination-related, comprehension-related, and motivational learning challenges.

For motivational challenges, though, we found regulation to be sensitive to the reported challenge as well—but in the opposite direction than we had expected. For one out of three motivational challenges, direct strategies were not more frequently used than other strategies. Since we proposed that there are more direct strategy types for regulating motivational than coordination- or comprehension-related challenges, this is even more surprising. When this baseline rate was controlled for, direct strategies were less significantly chosen for all three motivational challenges than expected under random distribution. Hence, learners who encounter motivational challenges seem to specifically avoid using motivational strategies. There are at least the three different explanations for this result: First, our theoretical considerations might have been wrong so that we failed to identify the right strategy types to directly address motivational challenges. Second, the matching we derived based on our theoretical understanding might be applicable but our participants were not able to select direct strategies and chose other strategies due to their lacking strategy knowledge or due to an inability to activate direct strategies here. Third, the social nature of the regulation situation might show differential effects of social desirability: According to this interpretation, it might be plausible that group members appreciate receiving help from another group member in understanding the learning content (= a comprehension-related challenge), or in negotiating the subsequent learning process (= coordination-related challenge). In case of motivational problems, however, it might be much less desirable to intervene as a group member. Trying to motivate a group member might be perceived as unwanted intrusion and attempt to change the attitudes towards the learning task and thus produce reactance within the group. For a less obvious and less socially conflict-loaded control of their group members' motivation, learners could have tried to indirectly motivate others through providing them assistance in learning the material.

Following Zimmerman and Moylan (2009), we supposed that if strategies were applied that match the learning challenge, then successful regulation should be likely. Regulation success, in turn, should result in satisfaction with the learning process. Therefore, we tested the association of the directness of strategy choice with satisfaction with the learning experience as an indication of criterion validity of our theoretical model on the fit between strategies and challenges. Controlling for random variations of satisfaction between groups and between individuals in groups, and for fixed and random effects of time, we found a significant link between the fit of strategy to challenge type and satisfaction with the learning experience only for coordination- and comprehension-related, but not for motivational challenges. Thus, we assume that our conceptualization of fit is actually predictive of regulation success measured by satisfaction rating, but differentially for specific challenge types. Additionally, this result mirrors the finding of the single case analysis of Järvelä et al. (2010).

## Limitations and conclusions

Of course, this study has limitations: First, we did not have a direct measure of regulation success. Using satisfaction as a more distal measure surely is influenced by more than just the directness of chosen strategies. Perhaps the relationship between strategy-challenge fit and learning outcomes would be stronger if learning outcomes would be assessed by aid of objective measures (e.g. standardized test measuring the acquired knowledge). Second, even though we investigated regulation in real groups, analyses were based on self-reported

strategy use. Video-based analysis of real group interactions would be valuable and could nicely complement our more subjective data. Third, we looked for the fit between challenge and strategy types on an aggregated level. If the individual open-answer descriptions of each challenge and the corresponding regulation strategies would be used to propose a new fit, the results might provide a more precise picture of an even larger amount of direct strategies. Thus, as before, we believe that the estimated relative frequencies were also underestimated in this study. As a consequence, we plan to determine the fit for each individual strategy to each individual challenge as a future step. Furthermore, we have proposed strategy fit only in reference to the biggest challenge that participants experienced. Students who appropriately coped with all challenges during their group meetings except for their biggest challenge were likely to have lowered the empirical relationship between fit and satisfaction.

An important educational implication of our study is the following: We saw that students do not regulate motivational challenges much with strategies that (theoretically) would fit. In case that this results proves reliable in further studies, students should be trained in effectively regulating motivational challenges in particular, to increase their use of direct strategies for such challenges. For individual learning situations, Eckerlein, Steuer, and Dresel (2018) already showed that effective motivational regulation can successfully be fostered through dedicated training programs. Also, scaffolds or scripts (Kollar, Fischer, & Slotta, 2007) may additionally prompt group learners to adjust their use of strategies to challenges for more effective and satisfying collaborative learning experiences.

## References

- Boekaerts, M. (1997). Self-regulated learning: A new concept embraced by researchers, policy makers, educators, teachers, and students. *Learning and Instruction, 7*, 161–186. [https://doi.org/10.1016/S0959-4752\(96\)00015-1](https://doi.org/10.1016/S0959-4752(96)00015-1)
- Eckerlein, N., Steuer, G. & Dresel, M. (2018). Training university students in motivational regulation. Manuscript submitted for publication.
- Engelschalk, T., Steuer, G., & Dresel, M. (2015). Wie spezifisch regulieren Studierende ihre Motivation bei unterschiedlichen Anlässen? *Zeitschrift Für Entwicklungspsychologie Und Pädagogische Psychologie, 47*, 14–23. <https://doi.org/10.1026/0049-8637/a000120>
- Freeman, K. A. (1996) Attitudes toward work in project groups as predictors of academic performance, *Small Group Research, 27*(2), 265–282.
- Järvelä, S., & Hadwin, A. F. (2013). New Frontiers: Regulating Learning in CSCL. *Educational Psychologist, 48*(1), 25–39. <https://doi.org/10.1080/00461520.2012.748006>
- Järvelä, S., & Järvenoja, H. (2011). Socially constructed self-regulated learning and motivation regulation in collaborative learning Groups. *Teachers College Record, 113*(2), 350–374.
- Järvelä, S., Järvenoja, H., & Veermans, M. (2008). Understanding dynamics of motivation in socially shared learning. *International Journal of Educational Research, 47*, 122–135.
- Järvelä, S., Volet, S., & Järvenoja, H. (2010). Research on motivation in collaborative learning: Moving beyond the cognitive-situative divide and combining individual and social processes. *Educational Psychologist, 45*(1), 15–27. <http://dx.doi.org/10.1080/00461520903433539>
- Järvenoja, H., Volet, S., & Järvelä, S. (2013). Regulation of emotions in socially challenging learning situations: an instrument to measure the adaptive and social nature of the regulation process. *Educational Psychology, 33*(1), 31–58. <https://doi.org/10.1080/01443410.2012.742334>
- Kollar, I., Fischer, F., & Slotta, J. D. (2007). Internal and external scripts in computer-supported collaborative inquiry learning. *Learning and Instruction, 17*, 708–721. <https://doi.org/10.1016/j.learninstruc.2007.09.021>
- Malmberg, J., Järvelä, S., Järvenoja, H., & Panadero, E. (2015). Promoting socially shared regulation of learning in CSCL: Progress of socially shared regulation among high- and low-performing groups. *Computers in Human Behavior, 52*, 562–572. <https://doi.org/10.1016/j.chb.2015.03.082>
- Mandl, H., & Friedrich, H. F. (Eds.). (2006). *Handbuch Lernstrategien*. Göttingen: Hogrefe.
- Singer, J., & Willett, J. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, NY: Oxford University Press.
- Winne, P. H., Jamieson-Noel, D., & Muis, K. Methodological issues and advances in researching tactics, strategies, and self-regulated learning. In *New Directions in Measures and Methods* (pp. 121–155). Elsevier Science Ltd.
- Zimmerman, B. J., & Moylan, A. R. (2009). Self-regulation: Where metacognition and motivation intersect. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *The educational psychology series. Handbook of metacognition in education* (pp. 299–315). New York, NY, US: Routledge/Taylor & Francis Group.

# Unpacking Socio-Metacognitive Sense-Making Patterns to Support Collaborative Discourse

Marcela Borge, Tugce Aldemir, and Yu Xia  
mborge@psu.edu, tfa5065@psu.edu, yzx64@psu.edu  
The Pennsylvania State University

**Abstract:** This study explores the dynamics between socio-metacognitive communication patterns and collaborative processes, as students engage in collaborative discussions about course concepts. Building upon a series of studies that aimed to design and validate an intervention to help students develop collaborative competencies at the group level, the study aims to map how socio-metacognitive sense-making patterns are associated with the collaboration quality, by comparing the patterns for low, medium, and high performing teams. Discussion and after-discussion reflection transcripts of 12 teams over five sessions were analyzed and assessed, using previously developed collaborative discourse rubric and sense-making coding construct. The results showed a significant correlation between frequency of sense-making acts and the quality of the collaborative discourse.

## Introduction

Collaborative competencies are essential sets of skills for society and it is vital to identify strategies and approaches to nurture them in learners. Unfortunately, many learners do not have the opportunities to develop collaborative competencies, which can lead to poor performance outcomes and undesirable group conflicts (Borge & Carroll, 2014; Fischer et al., 2013; Kozlowski & Ilgen, 2006). In addition to individual problems associated with underdeveloped collaborative skills, in CSCL environments, group-level issues such as social loafing, sucker effect (Salomon & Globerson, 1989), lack of non-verbal expressions, and time lag between the interactions can contribute to lower quality collaboration (Kreijns, Kirschner, & Jochems, 2003). This is why helping learners engage in higher quality collaborative processes has been a central concern for CSCL. As such, many CSCL researchers have attempted to help students improve collaborative skills, but few have recognized the complex and nested nature of collaboration (Borge & White, 2016; Borge, Ong Shiou, & Rosé, 2018; Baker et al., 2007; Stahl, 2006).

Prior research has indicated that without any guidance, support, or training, students tend to demonstrate dysfunctional group processes (Borge et al., 2018; Barron, 2003; Hogan, 1999; Webb & Palincsar, 1996). To address this problem, existing studies have designed and evaluated interventions to support group function, including (1) scripting collaborative interactions, (2) helping students improve metacognitive sense-making through reflecting upon individual and group performances and comparing them to the models of competence, (3) supporting students to develop self- and group-regulatory behaviors, and so forth. However, there still are issues that need further investigation, such as what kind of scripting is needed, how to balance the level of scripting, when to fade the external support, and how to help groups internalize the external support (Borge & White, 2016; Hogan, 1999; Kozlowski, Watola, Jensen, Kim, & Botero, 2009).

Building upon the critical role that self-regulation plays in individual's learning (Zimmerman & Schunk, 2001), Borge et al.'s (2018) suggested that metacognitive guidance and regulation at the group level can serve to help students optimize their collaborative experiences. They referred to collective awareness of and the collective ability to monitor and regulate the collaborative process as socio-metacognition. Borge et al.'s (2018) findings were promising, but also highlighted the need for further research on the development of socio-metacognitive expertise. Similarly, the scarcity of research on the interplay between socio-metacognitive process and collaborative discussion necessitates further exploration of their dynamics in CSCL context (Kwon, Liu, & Johnson, 2014; Rogat & Adams-Wiggins, 2015). This paper builds upon this existing literature and aims to extend what is known about collective regulation by identifying critical socio-metacognitive sense-making patterns in process-related dialogue acts in real-world collaborative learning contexts. Thus, we explored the dynamics between socio-metacognitive sense-making patterns and collaborative processes, as students engaged in collaborative discussions about course concepts.

## Theoretical framework

Our work is influenced by the theory of group cognition and thus recognizes collaboration as a form of nested cognition that entails intersubjective knowledge construction and collective sense-making that are situated at different levels, including individual, group, and community levels (Stahl, 2006). Collaborative activities offer

sophisticated learning opportunities that might not be available in individual learning activities, which are likely to help learners meet the need in our society for higher collaborative competencies and to succeed in collaborative teams. However, problems often arise in collaborative activities, especially in terms of the collective sense-making processes (Barron, 2003). The collaborative discourse that occur during group interactions has a large influence on the group performance and outcome (Kozlowski & Ilgen, 2006; Stahl, 2006). Thus, the examination of communication patterns can help evaluate the quality of collaborative discourse.

## Challenges for group regulation

Literature on self-regulation in individual learning has identified problems that pose barriers for learners to successfully regulate their own learning, such as not being aware of learning problems, misdiagnosing them, etc. (Winne & Nesbit, 2009). In collaborative contexts, these individual regulation problems arise alongside group regulation problems, where many teams fail to identify or accurately assess problematic processes and devise remediation strategies. There are also social and emotional difficulties involved in communicating problems and regulating group behaviors (Järvelä, & Hadwin, 2013). In each stage of regulating collaborative process, there is a requirement on building shared understanding through synthesizing information and idea negotiation (Stahl, 2006). We argue that in order for teams to engage in high-quality collaborative activities, they need to develop socio-metacognitive expertise: the knowledge of and ability to monitor and regulate collective cognitive processes.

## Socio-metacognitive sense-making

As collective literature argues, engaging in high quality collaborative discussion is a hard but crucial skill to develop. Despite multiple complexities and interrelated variables to consider, research also suggested that learners can improve the quality of their collaborative activities over time, if they learn how to regulate their collaborative process (Borge et al., 2018; Kozlowski et al., 2009).

Collaborative interactions are the externalized forms of collective thinking, and thus, how teams collectively make sense of, monitor, and regulate these interactions play a central role in collaborative process. Recognizing the role of group regulation in collaborative activities, recent research efforts have been directed toward developing technological tools to support group regulation. As one of the pioneers of these efforts, Järvelä and Hadwin (2013) developed a technological tool to help learners develop awareness and planning strategies for their collaborative activities. However, developing awareness and planning is not sufficient for the regulation process; learners need to develop an understanding for how a high-quality collaborative discussion should be like, compare their process to a model, identify problems in their process, collectively develop or choose appropriate remedial strategies to solve problems, and take action (Nesbit, 2012; Winnie, & Nesbit, 2009). In group contexts, these steps demand both individual and collective efforts, attention, and time, which might be the reason why teams tend to neglect their collaborative process while reserving all the attention to make sense of the content (Kerr & Tindale, 2004). Collective sense-making of the collaborative process requires teams to monitor and reflect on their collaborative activity (Nesbit, 2012). However, research suggested that individuals do not perform well at asking and addressing these questions, and situation gets even intensified when sense-making moves from individual to group cognition (Gabelica et al., 2014).

In our previous work, we addressed problems associated with collective regulation and developed a theoretically informed technological intervention to help students develop their socio-metacognitive expertise. We identified the communication patterns associated with high quality collaborative discussion, proposed two core capacities for collaborative sense-making, and listed concrete patterns of communication associated with more or less optimal collaborative sense-making processes. We then helped students monitor and regulate their collaborative processes by guiding and constraining how they get prepared for discussion and how they engaged in socio-metacognitive sense-making and regulation after the discussion. The intervention succeeded in getting teams to improve the quality of their collaborative activity over time. However, we did not fully examine the dynamics between socio-metacognitive competence, collective sense-making, and the quality of the collaborative discussion. The collection of research on socio-metacognitive sense-making suggests that it is critical to identify how teams engage in socio-metacognitive sense-making and regulation of both course content and team process discourse, and how these interconnected processes may impact each other. To address this need, we aim to identify socio-metacognitive sense-making patterns in collaborative discussions and thus examined how these patterns are associated with team collaborative performance. Our research questions were:

(RQ1) What patterns of socio-metacognitive sense-making (SMS) talk do teams engage in when unpacking course content and thinking about their own discussion processes?

(RQ2) What are the differences in these patterns between low, medium, and high performing teams?

## Methods



## Course context and participants

The study was conducted in an online 16-week undergraduate course designed to introduce students to information science concepts. As part of the course, students were expected to engage in collaborative reasoning practices and discussion activities. Developing collaborative discussion skills was one main goal of the course. Participants were 34 online students who enrolled in the course (11 females, 33.3%; 22 males, 66.7%). Students' ages ranged from 25 to 44, and the majority of students were part-time students with full-time jobs.

## Procedure

Students were assigned to 12 teams of three based on when they were available to meet. Due to two students dropping the class, two teams ended up as dyads. These teams were required to meet synchronously for five sessions to collectively make sense of course concepts. They met every other week for ten weeks. As a pre-discussion activity, the students were required to read the weekly readings and write an individual reflection in response to four higher-order questions about the readings. Then, they were asked to set a meeting time with their teammates to synchronously discuss the questions and readings. Each discussion session was about 90 minutes: 60-minute main discussion, 15-minute individual assessment of team discussion, and 15-minute collective planning discussion. These activities counted towards 25% of students' grades. For the 15-minute individual assessment of the team discussion that followed the 60-minute main discussion, students were provided with a collaborative process rubric detailing how to assess discussion quality, guides containing goals for collaboration, problems that interfered with good collaboration, and strategies for improving collaborative processes. The discussions were held on a computer supported collaborative discussion environment, and saved automatically in the system. After each individual scored their team's collective processes, the entire team was responsible for completing a collective planning session, where they discussed their scores and process weaknesses they identified during individual reflection for the purpose of collectively diagnosing problems and planning out strategies the team could use to improve in future sessions.

## Research design and analysis

We implemented explanatory mixed methods (Creswell, 2015). The main discussions, individual reflections, and collective planning sessions were collected and analyzed for 12 teams across five sessions, following the verbal analyses guidelines offered by Chi (1997). We coded a total of 12,755 utterances, 10,764 from content-based discussions and 1,991 from team planning sessions. All ethical guidelines were followed in collecting, analyzing, and reporting the study.

## Evaluating teams' discourse quality

Building upon previous theoretical frameworks (Borge et al., 2018), teams' discourse quality, when teams work to collectively understand course content, is defined as the extent to which team members provide evidence of engaging in communication patterns associated with high-quality information synthesis and knowledge negotiation (see Table 1). The quality of teams' discussions was assessed by a research assistant with two years of communication analysis training, using a rubric developed by Borge et al., (2018), which measures two core capacities each with three categories of behavior (see Table 1). To score each item, the entire transcript is

Table 1: Summary of core capacities and categories of behavior associated with high quality collaborative communication behaviors from Borge et al. (2018).\*

| <b>Core Capacities</b>   | <b>Categories of Behavior</b>         | <b>What is Examined in the Transcript</b>  |
|--------------------------|---------------------------------------|--|
| Information<br>Synthesis | Verbal Equity                         | To what extent team members contributed equally.   |
|                          | Developing Joint Understanding        | To what extent team members make an effort to ensure that they fully understand the ideas. |
|                          | Joint Idea Building                   | To what extent team members elaborate on others' contributions.                            |
| Knowledge<br>Negotiation | Exploration of Different Perspectives | To what extent teams present and discuss alternative perspectives.                         |
|                          | Quality of Claims                     | To what extent teams provide logical and fact-based evidence to their claim.               |
|                          | Norms of Evaluation                   | To what extent teams adhere to social norms.   |

*Note:* \* Each score ranging from 1 to 5 outlines a set of guidelines indicating what each score means for each category. For example, in Quality of claims, a score of five means "There are at least two examples where claims are supported by references to course readings or online content AND at least one example of weighing of options or examination of different perspectives." (Borge et al., 2018).

examined for specific discourse quality markers (see top of figure 1 for example) and these markers are used as a means to provide evidence for a score from 1 to 5 (see bottom of Figure 1). Scores for the six categories were summed to a single collaborative discussion quality score for each discussion. Once all the five main discussion scores were identified, they were averaged to produce the average collaboration performance of each team. The quality of the collaborative discussion was assessed using only the main discussion transcripts. 20% of the data were double coded by two trained students with extensive communication analysis experience. Significant agreement was reached,  $r = .86$ ,  $p < .001$ ; Kappa = .64,  $p < .001$ .

| Session | Turn | User ID | Entry   | Codes         |
|---------|------|---------|---|---------------|
| 1       | 29   | 86      | He tries to balance his views with a pro for privacy violations, asserting that it may be used for national security. Do you agree with this implication? | DJU-1, JIB- 1 |
| 1       | 31   | 85      | I apologize for the delay, I've been rereading your post, I'm not sure I understand what you mean. The author is pro privacy violation?                   | DJU- 1        |

| Criteria                       | Score | Evidence  |
|--------------------------------|-------|---|
| Joint Idea Building            | 4     | I could not locate more than one instance [of JIB] where the participants extended or supported the original claim with additional information. The participants seemed to either skip to another subject very quickly or discuss an opinion over one or two turns.   |
| Developing Joint Understanding | 5     | There are at least two instances where one participant made an effort to ensure s/he understood what the previous participant said. They used the following strategies: (DJU-1) Rewords another member's ideas to make sure s/he understands it; (DJU-2) Asks another member to explain an idea by elaborating further. |

Figure 2. Two screenshots depicting how lines of transcript are examined for specific markers for developing joint understanding (DJU-1 represents first instance of DJU) and joint idea building (JIB-1 represents first instance of JIB) (top) and how these markers are used as evidence for discourse quality scores (bottom).

### Analysis of socio-metacognitive sense-making (SMS) talk

The transcripts of main discussions and after-discussion collective reflection and planning sessions were analyzed for 12 teams across five sessions, for a total of 60 analyzed discussions. Socio-metacognitive sense-making is a specific type of process talk where students think about their collaborative processes to try to understand or modify them. To identify SMS talk, main discussions and reflection discussions were segmented into chat turns; then each turn was coded as process (P), content (C), or other (O). Inter-rater reliability for 20% initial coding was Kappa = .79,  $p < .001$ . The total frequency of P acts varied for each team. To compare teams' SMS behaviors, we calculated percentages of SMS talk out of P acts.

We used two versions of a socio-metacognitive sense-making coding scheme originally developed by Borge et al. (2018): the original version was used for reflections about the processes that occurred during their content-based discussion and a second version for the content-based discussion itself. Both original and modified rubrics are presented in Table 2. The original version was designed to code talk that occurred during reflection sessions to identify the extent to which teams engaged in socio-metacognitive sense-making activities during the reflective activity. It included process reporting, monitoring, reflecting, planning, and revising. Two trained coders

Table 2: Examples of SMS patterns by the participants

| SMS Pattern             | Rubric Version  | Example   |
|-------------------------|-----------------|---|
| Reporting (RP)          | Both rubrics    | <i>"We sure covered a lot."</i> (MD)<br>A student reports her/his opinion about the collaboration quality without referring to concrete events or patterns from discussion or justifying her/his judgment.)   |
| Process Monitoring (MO) | Both rubrics    | <i>"and we also evaluated trade-offs for some while comparing implications."</i> (MD)<br>(A student demonstrates evidence of paying attention to the ongoing collective process by pointing out a specific activity/requirement that the team has done or needs to do.)   |
| Process Reflection (RF) | Both rubrics    | <i>"I think the problem is that we read 2 different things."</i> (MD)<br>(A student demonstrates evidence of reflecting on their ongoing collaborative discussion by pointing out a reason for why s/he thinks that particular incident happened.)  |
| Process Planning (PL)   | Original rubric | <i>"To begin, we definitely need to work on time management. Our communication skills are sufficient when it comes to the subject matter, but we definitely need to get tasks done with a sense of urgency."</i> (AR) (A student demonstrates evidence of planning by unpacking the problem and proposing new goals.) |
| Regulation (R)          | Modified rubric | <i>"Let me play devil's advocate, since we need to think about the other side of the coin."</i> (MD)  |

(A student demonstrates evidence of regulating ongoing activity reminding teammates about previously set process goals.)

Note: MD=Main discussion, AR=After-discussion reflection.

coded 23% of the total reflective discussion data using the original construct, with Kappa = .806;  $p < .001$ . The coders discussed and resolved disagreements and then one coder re-coded all the reflection data. A second version of the coding construct was created to code SMS talk that occurred during the content-based discussion. This new version included all the previous forms of talk plus a new category: regulation talk. Regulation talk identified socio-metacognitive strategies, including moves such as proposing or using a discussion strategy from guides we provided. Inter-rater reliability was checked on this new version on 24% data with Kappa = .725;  $p < .001$ . Two coders discussed and resolved disagreements and then one coder re-coded all the main discussion data.

### Selection of high, medium, and low performing teams

Average collaborative discussion quality scores were used to assess performance for each team (see Table 3). High, medium, and low performing categories each have four teams.

Table 3: Mean discussion quality scores of the teams (Lowest to Highest)

|    | Low Quality Discourse Teams |       |       |       | Medium Quality Discourse Teams |       |       |       | High Quality Discourse Teams |       |       |       |
|----|-----------------------------|-------|-------|-------|--------------------------------|-------|-------|-------|------------------------------|-------|-------|-------|
|    | T9                          | T7    | T1    | T4    | T3                             | T6    | T8    | T12   | T5                           | T10   | T11   | T2    |
| M  | 19.80                       | 20.20 | 21.40 | 21.80 | 23.40                          | 23.40 | 23.40 | 23.80 | 24.00                        | 24.40 | 24.60 | 25.40 |
| SD | 3.56                        | 4.15  | 3.36  | 3.42  | 2.88                           | 2.51  | 3.97  | 2.59  | 3.32                         | 1.52  | 3.65  | 3.21  |

### Comparing team process interactions

We conducted one-way ANOVA to determine whether there was a significant difference in the average means of the amount of socio-metacognitive sense-making talk that occurred in low, medium, and high performing teams. For that, Levene's test for homogeneity of variances was found to be protected for the total percentage of SMS acts ( $F(2,9)= 1.23, p=.34$ ). Shapiro-Wilk test was performed to ensure normality, and no significant value was identified; therefore, the normality assumption was met to perform ANOVA.

## Findings

### (RQ1) Patterns of socio-metacognitive sense-making (SMS) talk

On average, each team created 263.58 process-related turns over five discussion sessions,  $SD=120.75$ ,  $Min=133$ ,  $Max=521$ . Over half of these process-related turns, 54.54% ( $SD=13.03$ ), were coded as SMS talk. Those that were not coded as SMS focused on sharing information about how they were using technology or other forms of social-off-task talk. All teams engaged in RP, MO, RF, and PL/R acts in at least one discussion session, but only 7 out of 12 teams demonstrated RV act (Table 4).

Table 4: The mean percentages of teams' SMS patterns

|                                    | Main Discussions |       | After-Discussion Reflections |       | Total     |      |
|------------------------------------|------------------|-------|------------------------------|-------|-----------|------|
|                                    | M (%) *          | SD    | M (%) **                     | SD    | M (%) *** | SD   |
| Reporting (RP)                     | 2.65             | 4.96  | 18.37                        | 14.65 | 11.52     | 5.06 |
| Process Monitoring (MO)            | 12.27            | 4.77  | 8.45                         | 7.34  | 10.39     | 4.76 |
| Process Reflection (RF)            | 0.19             | 0.45  | 5.44                         | 2.63  | 3.45      | 1.66 |
| Process Planning/Regulation (PL/R) | 43.93            | 11.05 | 15.14                        | 11.96 | 28.64     | 9.71 |
| Process Revising (RV)              | 0.58             | 1.15  | 0.45                         | 0.67  | 0.54      | 0.62 |

Notes: \* Mean frequency percentage out of the total P frequency in main discussions for each SMS pattern

\*\* Mean frequency percentage out of the total P frequency in after-discussion reflections for each SMS pattern

\*\*\* Mean frequency percentage out of the total P frequency for each SMS pattern

The findings suggested that the teams engaged in PL/R act the most. In the main discussion these acts occur as regulation acts focused on proposing a discussion strategy or new direction for conversation, whereas in the after-discussion reflections these acts are planning acts focused on identifying a strength or weakness, or proposing and evaluating goals and strategies for future discussions. As shown in table 4, the high frequency of PL/R acts are due to the high number of regulation acts that occurred in the main discussions. For example, during the second discussion, a team was discussing whether all members understood the course content. One team member, Bill, stated that the content was new to them so they did not have full understanding of it. Upon hearing that Bill had specific questions, another member, Juan, said: "... we had some questions there, Bill, you can start" (R act in Main Discussion; Team 2, Session 2).

Figuring out how to discuss topics deeply, while keeping to an agreed upon time limit was a common topic discussed during the after-discussion reflection sessions. Team 2 discussed this topic as part of their reflection. Bill said, “Although we got very in depth, I feel like we could get just as in depth if we focus the conversation more and have strict time framing.” Another member, Jill, added, “We need to figure out how to keep to the timeframe without cutting off something important” (PL Act, After-Discussion Reflection; Team 2, Session 2,). Juan then proposed a suggestion for improvement, “We can do anything! we will just have to weigh our questions maybe before we start the conversation and start there as a base.” Bill responded by saying, “Right maybe ask the longer ones first” (PL Act, After-Discussion Reflection).

## (RQ2) Differences between low, medium, and high-quality discourse teams

Looking at the relationship between SMS talks and team performance, when ranking teams by SMS acts, we saw that the frequency of SMS acts was closely tied to the quality of collaborative discourse (see table 5).

Table 5: Frequencies of teams’ collaboration patterns (SMS Lowest to Highest for Each Performance Category)

|      | Low Quality Discourse Teams |       |       |       | Medium Quality Discourse Teams |       |       |       | High Quality Discourse Teams |       |       |       |
|------|-----------------------------|-------|-------|-------|--------------------------------|-------|-------|-------|------------------------------|-------|-------|-------|
|      | T4                          | T7    | T9    | T1    | T6                             | T8    | T12   | T3    | T11                          | T5    | T10   | T2    |
| P    | 348                         | 365   | 196   | 217   | 142                            | 133   | 247   | 200   | 216                          | 521   | 166   | 412   |
| SMS* | 25.57                       | 41.10 | 43.88 | 50.23 | 47.89                          | 57.89 | 61.54 | 67.00 | 60.19                        | 65.45 | 66.27 | 67.48 |

Notes: \* Total frequency percentage out of the total P frequency for each team  
P=Process-related talk, SMS=Socio-metacognitive sense-making talk, T=Teams

Comparing values on Table 5 with those on Table 3 (teams’ mean discussion quality scores), we observed that low quality discourse teams tended to engage in least SMS acts while high quality discourse teams tended to engage in most SMS acts. However, as illustrated in Table 5, there were some exceptions. A Pearson correlation analysis on SMS percentage means and mean performances showed a strong positive correlation ( $r = .749, n = 12, p = .005$ ). Analysis of variance showed a significant difference in the SMS percentage means between different quality discourse teams,  $F(2,9)=10.66, p=.00$ . Tukey HSD indicated a significant difference in SMS percentage means between low ( $M=40.20, SD=10.47$ ) and high-quality discourse teams ( $M=64.85, SD=3.22$ ), and between low ( $M=40.20, SD=10.47$ ) and medium quality discourse teams ( $M=58.58, SD=8.05$ ). No significant difference was found between medium and high-quality discourse teams.

We conducted one-way ANOVA to determine whether there is a significant difference in the five SMS patterns ( $n=5$ ) between the low, medium, and high-quality discourse teams; no significant differences were found. However, descriptive analysis of the percentage values between teams yielded interesting trends. The average percentage of PL/R act increased from low ( $M=19.76, SD=5.12$ ) to medium performing teams ( $M=32.09, SD=9.92$ ), and from medium to high performing teams ( $M=34.09, SD=7.77$ ). In RV act: low-performing teams engaged less ( $M=0.27, SD=0.31$ ) than medium-performing teams ( $M=0.52, SD=0.37$ ), who engaged less than high-performing teams ( $M=0.83, SD=0.98$ ). The average percentage of RF act in low-performing teams was smaller ( $M=2.86, SD=1.21$ ) than that of medium-performing teams ( $M=3.62, SD=2.65$ ), which was smaller than that of high-performing teams ( $M=3.87, SD=0.91$ ). Figure 2 below summarizes the distribution of the SMS acts by performance categories.

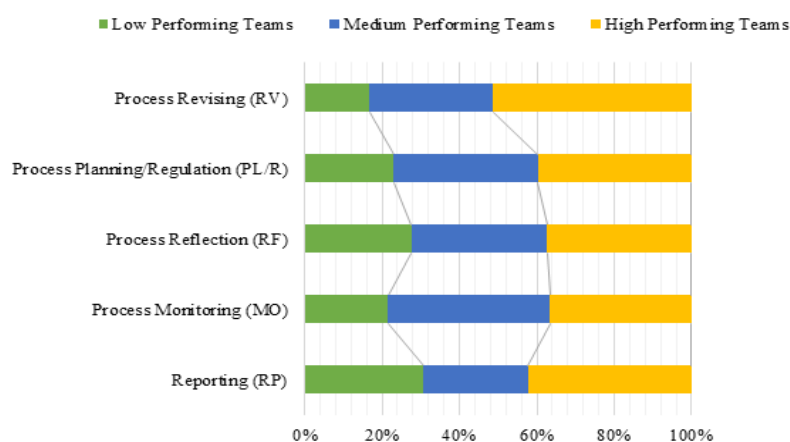


Figure 2. SMS percentage means of the teams by performance categories.

## Discussion

In this paper we examined teams' sense-making patterns in the main content-based discussions and during after-discussion reflections, where teams made sense of their collaborative process. We also examined the relationship between sense-making patterns and collaborative discourse quality. Our findings showed that all the teams engaged in collective socio-metacognitive sensemaking (SMS) talk during content-based discussions and after-discussion reflections. Further supporting the design of the original intervention, which aimed at pushing students to engage in SMS talk as part of reflections for the purpose of getting them to figure out regulate activity during content-based talk and then actually regulate activity. We also found a strong positive correlation between the frequency of socio-metacognitive sense-making talk and the quality of collaborative activity.

These findings add to the growing body of research that indicate that supporting collective regulation of group interactions may be the key to enhancing the quality of collaborative processes (Kwon, Liu, & Johnson, 2014; Rogat & Adams-Wiggins, 2015). Prior research has suggested that students do not have the ability to monitor and regulate individual and collaborative activities (Borge & White, 2016; Kwon, Liu, & Johnson, 2014; Gabelica et al., 2014; Winne & Nesbit, 2009), and if not provided with sufficient amount of guidance, they will likely to develop dysfunctional collaborative habits (Borge et al., 2018; Kozlowski & Ilgen, 2006; Webb & Palincsar, 1996). The current study suggests that guiding students to develop sense-making skills to monitor and reflect their ongoing/past collaborative processes can help teams to developing sophisticated collaborative competencies. What is more, this work suggests it is possible to enhance collaborative processes without over-scripting collaboration as it occurs or creating inauthentic collaborative environments (Dillenbourg, 2002; Fischer et al., 2013). As such, this paper also extends what is known about the impacts of socio-metacognition on teams' ability to improve the quality of collaborative discussion processes.

Our findings imply that scripting socio-metacognitive sense-making activities before and after collaborative activity can help teams to regulate that ongoing collaborative activity, thereby developing socio-metacognitive competence and optimizing their ongoing collaborative activities. However, it also underlines the need for a more blended approach to scripting that supports collaborative process both during and after they occur. Nonetheless, it is crucial to note that further research is needed to support these findings and to explore how after-discussion reflection impact students' sense-making of their collaborative process, their socio-metacognitive development, and collaborative process improvement.

One limitation of this study is the low number of the groups for each quality discourse category. Further studies are needed to see if the patterns observed in the study remain once the number of observations increases. Furthermore, the focus of this study did not fully address impacts of emotions on social interactions, which were also prevalent forms of process-based talk. Building on our theoretical framework, collaboration emerges as a product of interactions among individuals (Stahl, 2006), which entails the exchange of ideas, knowledge, emotions, and feelings (Järvenoja & Järvelä, 2009). Thus, further research is needed to examine the dynamics between socio-metacognitive expertise, socio-emotional interactions, and the quality of collaborative discourse.

## References

- Baker, M., Andriessen, J., Lund, K., van Amelsvoort, M., & Quignard, M. (2007). Rainbow: A framework for analyzing computer-mediated pedagogical debates. *International Journal of Computer-Supported Collaborative Learning*, 2(2), 315-357.
- Barron, B. (2003). When smart groups fail. *The Journal of the Learning Sciences*, 12(3), 307-359.
- Borge, M., & Carroll, J. M. (2014, November). Verbal equity, cognitive specialization, and performance. In *Proceedings of the 18th International Conference on Supporting Group Work* (pp. 215-225). ACM.
- Borge, M., & White, B. (2016). Toward the development of socio-metacognitive expertise: An approach to developing collaborative competence. *Cognition and Instruction*, 34(4), 323-360.
- Borge, M., Ong, Y. S., & Rosé, C. P. (2018). Learning to monitor and regulate collective thinking processes. *International Journal of Computer-Supported Collaborative Learning*, 13(1), 61-92.
- Bruner, J. (1999). Postscript: Some reflections on education research. In E. C. Lagemann & L. S. Shulman (Eds.), *Issues in education research: Problems and possibilities* (pp. 399-409). San Francisco: Jossey-Bass Publishers.
- Chi, M. T. H. (1997). Quantifying Qualitative Analyses of Verbal Data: A Practical Guide. *Journal of the Learning Sciences*, 6(3), 271-315.
- Dillenbourg, P. (2002). Over-scripting CSCL: The risks of blending collaborative learning with instructional design. In P. A. Kirschner (Eds.), *Three worlds of CSCL. Can we support CSCL* (pp. 61-91). Heerlen: Open Universiteit Nederland.

- Fischer, F., Kollar, I., Stegmann, K., Wecker, C., Zottman, J., & Weinberger, A. (2013). Collaboration scripts in computer supported collaborative learning, In C. Hmelo-Silver, C. Chinn, C. Chan, & A. O'Donnell (Eds.), *The International handbook of collaborative learning* (pp. 403-419). New York, NY: Routledge.
- Gabelica, C., van den Bossche, P., De Maeyer, S., Segers, M., & Gijsselaers, W. (2014). The effect of team feedback and guided reflexivity on team performance change. *Learning and Instruction, 34*, 86-96.
- Hogan, K. (1999). Thinking aloud together: A test of an intervention to foster students' collaborative scientific reasoning. *Journal of Research in Science Teaching, 36*(10), 1085-1109.
- Järvelä, S., & Hadwin, A. F. (2013). New frontiers: Regulating learning in CSCL. *Educational Psychologist, 48*(1), 25-39.
- Järvenoja, H., & Järvelä, S. (2009). Emotion control in collaborative learning situations: Do students regulate emotions evoked by social challenges. *British Journal of Educational Psychology, 79*(3), 463-481.
- Kerr, N., & Tindale, S. (2004). Group performance and decision making. *Annual Review of Psychology, 55*(1), 623-655.
- Kozlowski, S., & Ilgen, D. (2006). Enhancing the Effectiveness of Work Groups and Teams. *Psychological Science in the Public Interest, 7*(3), 77-124.
- Kozlowski, S. W., Watola, D. J., Jensen, J. M., Kim, B. H., & Botero, I. C. (2009). Developing adaptive teams: A theory of dynamic team leadership. In S. Eduardo, G. F. Goodwin, & C. S. Burke (Eds.), *Team effectiveness in complex organizations: Cross-disciplinary perspectives and approaches* (pp.113-155). New York, NY: Routledge.
- Kreijns, K., Kirschner, P. A., & Jochems, W. (2003). Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: A review of the research. *Computers in Human Behavior, 19*(3), 335-353.
- Creswell, J. (2015). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. New York, NY: Pearson.
- Kwon, K., Liu, Y. H., & Johnson, L. P. (2014). Group Regulation and Social-Emotional Interactions Observed in Computer Supported Collaborative Learning: Comparison Between Good vs. Poor Collaborators. *Computers & Education, 78*, 185-200.
- Nesbit, P. L. (2012). The Role of Self-Reflection, Emotional Management of Feedback, and Self-Regulation Processes in Self-Directed Leadership Development. *Human Resource Development Review, 11*(2), 203-226.
- Rogat, T. K., & Adams-Wiggins, K. R. (2015). Interrelation between Regulatory and Socioemotional Processes within Collaborative Groups Characterized by Facilitative and Directive Other-Regulation. *Computers in Human Behavior, 52*, 589-600.
- Salomon, G., & Globerson, T. (1989). When teams do not function the way they ought to. *International Journal of Educational Research, 13*, 89-99.
- Stahl, G. (2006). *Group cognition: Computer support for building collaborative knowledge*. Cambridge, MA: MIT Press.
- Webb, N. M., & Palincsar, A. S. (1996). Group processes in the classroom. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 841-873). New York, NY: Macmillan.
- Winne, P. H., & Nesbit, J. C. (2009). Supporting self-regulated learning with cognitive tools. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 259-277). New York, NY: Routledge.
- Zimmerman, B. J., & Schunk, D. H. (2001). *Self-regulated learning and academic achievement: Theoretical perspectives* (2<sup>nd</sup> Eds.). Routledge.

# The Binary Replicate Test: Determining the Sensitivity of CSCL Models to Coding Error

Brendan R. Eagan, University of Wisconsin–Madison, beagan@wisc.edu

Zachari Swiecki, University of Wisconsin–Madison, swiecki@wisc.edu

Cayley Farrell, University of Wisconsin–Madison, cefarrell@wisc.edu

David Williamson Shaffer, University of Wisconsin–Madison and Aalborg University, Copenhagen,  
dws@education.wisc.edu

**Abstract:** The process of labeling, categorizing, or otherwise annotating data—or *coding* in the computer-supported collaborative learning (CSCL) literature—is a fundamental process in CSCL research. It is the process by which researchers identify salient properties about segments of CSCL data: what they are, what they contain, or what they mean. Coding, like all processes in research, is subject to error. To reduce the potential impact of coding error, CSCL researchers typically measure inter-rater reliability (IRR). However, there is no extant method to determine what level of IRR would invalidate a CSCL result or model. One way of assessing the potential impact of such inaccuracies is by conducting *sensitivity analyses*, which measure the *level of error* that would need to be present in the data to invalidate a given inference. This paper introduces a new method for conducting sensitivity analyses in CSCL: the Binary Replicate Test.

## Introduction

Inaccuracies in measuring collaborative learning in *computer-supported collaborative learning* (CSCL) environments can result in poor learning outcomes, inappropriate or ineffective pedagogical responses, misguided design decisions, or misallocation of scarce resources. In order to effectively *control* for the impact of measurement error, researchers need to be able to *measure* it. One general approach to assessing the impact of inaccuracies in data is through *sensitivity analyses*, which measure the *level of error* that would need to be present in the data to invalidate a given inference: that is, the extent to which the data could be altered until the original result becomes invalid (Frank & Min, 2007) <sup>(1)</sup>. Here, we develop and test a method to apply this general strategy to the specific case of *coding error*.

In CSCL research, one of the largest sources of error involves *coding*. At its most basic level, coding is the process used to classify data: to assert that it should be labeled with a given construct or thought of as containing a given property (Chi, 1997; Shaffer, 2017). There are, of course, many sources of coding error. For example, hand coding performed by qualitative researchers is subject to individual bias, misinterpretation, or even just human fatigue—ask any researcher rating their 250<sup>th</sup> excerpt of data! Similarly, automated coding processes introduce some level of error, no matter how good the algorithm or approach. To reduce the potential impact of coding error, CSCL researchers typically measure inter-rater reliability (IRR) (Eagan, et. al. 2017; Hammer & Berland, 2014). While the common method of establishing IRR in CSCL research requires some methodological refinement (Eagan, et. al. 2017; Hammer & Berland, 2014; Shaffer, 2017), even when applied correctly, there is no extant method to determine what level of IRR (and thus what level of coding error) would invalidate a result or model.

In this study, we propose and provide an example of a method for conducting a Monte Carlo rejective test, a form of null hypothesis significance test, that empirically measures the impact of coding error on the results of any CSCL analysis. We consider the specific case of *binary coding*—the assertion of the presence or absence of a given construct for some observation—because this is a commonly-used approach to coding in CSCL research, and we refer to the proposed technique as a Binary Replicate Test (BRT). And in this initial paper on the BRT method, we apply the BRT approach using one specific modeling tool and one specific set of CSCL data. Our results suggest, however, that the BRT method has the potential to empirically determine thresholds for IRR measurement in CSCL analyses more generally.

## Theory

### Coding error

Binary coding is the process of assigning one of two possible values (typically 1 or 0) to pieces of data, reflecting the presence or absence of some property of interest in the data: for example, whether a student is asking a question in a particular turn or talk (1) or not (0). It follows that there are two types of error in binary coding. *Type I errors*

(false positives) are when an observation *does not* contain a given construct but is *coded* for the construct—for example, a turn of talk coded with a 1, indicating that the student is asking a question, is a Type I error if the student is, in fact, not asking a question. *Type II errors* (false negatives) are when an observation *does* contain a construct but is *not coded* for the construct—for example, if a turn of talk is coded with a 0, indicating the student was not asking a question, when she actually was.

In most CSCL settings, however, determining Type I and Type II errors is complex, because there is not necessarily a clear ground truth: two observers may see the same event in the data and code it differently. Researchers typically use measures of IRR, such as Cohen's kappa, to determine the overall rate of agreement—that is, some combined level of Type I and Type II errors between two raters, taking one rater as the standard to which the second rater is compared. The assumption—often implicit—in using IRR in this way is that the ground truth lies somewhere between the two sequences of 1s and 0s that each coder assigns to the data <sup>(2)</sup>.

Put another way, the IRR between two raters establishes an *upper bound* on the distance that either rater can be from ground truth—assuming that both raters are conscientious and understand the concept being coded, of course. Therefore, to assess the sensitivity of a given model to coding error, we need to determine the minimum level of an IRR statistic for which the model remains valid: that is, some level at which we can change 1s to 0s and 0s to 1s without altering the result of a given analysis conducted on the coded data.

In the case of IRR, guidance for establishing such thresholds is scarce. Many statistics have no commonly accepted threshold for what constitutes an acceptable level of agreement. For example, researchers refer to levels of Precision, Recall, or F statistics as being “high” without ever defining what constitutes a “high” level for that statistic (Schwartz, & Ostendorf, 2005). For other metrics, there are commonly used levels of acceptable agreement that have no statistical or inferential foundation. For example, Viera & Garrett (2005) suggested that “substantial agreement” for Cohen's kappa is above 0.61, and that has been used as a standard in many fields (Andrews, Leonard, Colgrove, & Kalinowski, 2011; Goh et al., 2008; Lasorsa, Lewis, & Holton, 2012), despite the fact that neither they, nor Cohen, provided a justification for that choice.

## Thresholds of significance

This use of IRR, is, of course, a form of sensitivity analysis. Sensitivity analyses in general quantify the conditions under which the validity of a result could be called into question by establishing a *threshold*: the point at which a particular level of error would invalidate an inference based on that result. Such analyses are often conducted using *Monte Carlo* (MC) studies (Egan et al. 2017; Harwell, 1992). MC studies assess the performance and reliability of statistical tests in educational and psychological research by creating an *empirical sampling distribution*. That is, they create a large number of simulated datasets and calculate a test statistic for each one. The performance of the test can then be evaluated under different assumptions about the properties of the population from which the samples are drawn, or based on variations on the sampling procedure itself.

For example, Egan and colleagues (2017) used MC studies to demonstrate the unacceptably high Type I error rates associated with using commonly applied methods for IRR measurement in CSCL, learning analytics, and related fields. Similarly, Harwell, Rubinstein, Hayes & Olds (1992) used MC studies to show that the Welch test, a variant of ANOVA, has inflated Type I error rates and lower power when used with non-normal distributions than with normal distributions. In the case of coding error, researchers can create simulated code sets by introducing different levels of error to an *actual* dataset to examine the performance of a model under different error conditions.

In this paper we conduct a sensitivity analysis on the results of one analytic technique used in CSCL and learning analytics, epistemic network analysis, but the BRT method is agnostic to the specific analytic approach.

## Model significance

Analytic models can be used to detect differences between groups, such as contrasting treatment and control, comparing different teams, highlighting instructor differences, and so on. In addition, analytic models can be used to assess learners at an individual level, which is useful for, among other things, providing feedback, grading, and selecting interventions. As a result, researchers need methods for establishing the accuracy of a learning analytic model for assessing both group differences and individual learners.

## MC rejective methods

Researchers can establish the robustness of analytic results using *rejective methods*. Using a criterion for rejecting a null hypothesis, such as employing p-value thresholds in hypothesis testing, is an example of a rejective method. Classically, when researchers observe a p-value under a critical value—often  $\alpha < 0.05$ —they reject the null hypothesis under the assumption that they would be doing so incorrectly 5% of the time if the null hypothesis were indeed true. While many rejective methods rely on a theoretical model, such as parametric tests using p-



values, others rely on empirical simulations to establish a criterion for rejecting a null hypothesis. For example, the MC rejective method Shaffer's rho generates an empirical null hypothesis distribution used to determine the likelihood of seeing a specific IRR measurement given the null hypothesis—that the actual IRR was below a threshold of interest (Shaffer, 2017). Similarly, an MC rejective method can be used to test a given analytical model's robustness to coding error when measuring group or individual differences.

## ENA

One technique that has been frequently used to analyze coded data produced in CSCL environments is *epistemic network analysis* (ENA) (see Shaffer, 2017, and references). To do so, ENA takes interaction data coded for elements of complex thinking, collaboration, learning, or any other phenomena of interest and creates weighted network models of the connections between those elements for individuals or groups.

As with any learning analytic technique applied to coded data, errors in coding can significantly affect statistical results and interpretations. In what follows, we present a study on the impact of coding error on one particular ENA result. However, we emphasize that our goal is less to show a method for conducting sensitivity analyses on one particular modeling technique than it is to present this as an example of a method that can be used to conduct sensitivity analyses on any CSCL or learning analytic result generated from coded data.

Specifically, we ask: ***How much coding error can be introduced to a dataset before we are no longer confident in the statistical significance (at the group level) and accuracy (at the individual level) of a given CSCL result?*** We address this research question by using a MC rejective method to examine the impact of coding error on one validated ENA result.

## Methods

### Gold-Standard ENA model

*General approach.* Our general approach to developing and testing a methodology for conducting sensitivity analyses was to create a “Gold-Standard” result: that is, an original, statistically significant result using coded data whose sensitivity we wanted to test. To do so we created an ENA model using a well-studied dataset (Arastoopour et al. 2016; Chesler et al., 2015). We then evaluated the effects of making perturbations to the coded dataset upon which the Gold-Standard result was based. Specifically, we measured the sensitivity of the ENA model to coding error by randomly introducing coding error into the dataset used to create the Gold-Standard model, and then looked to see the level of error at which the original result was no longer statistically significant or accurate at the individual level.

*Data source.* To create our Gold-Standard result, we analyzed the discourse of students in the engineering virtual internship *RescuShell* (Chesler et al., 2015). In *RescuShell*, students work in project teams to design robotic exoskeletons for use by rescue workers in disaster situations. The Gold-Standard model has two conditions: (a) students in the first condition (relative novices, hereafter referred to as Novices) were participating in an engineering virtual internship for the first time; (b) students in the second condition (relative experts, hereafter referred to as Experts) had previously participated in a different engineering virtual internship.

*Model.* Previous research (Arastoopour et al. 2016) found differences in the discourse patterns of Novices and Experts that were both statistically and interpretively significant (that is, meaningful). Chesler and colleagues (2015) found that Novices made connections mostly among basic skills and knowledge while Experts made additional connections with epistemological elements of engineering which are indicative of complex problem solving in engineering. To create the Gold-Standard model we replicated this previous research by applying Epistemic Network Analysis (ENA; Shaffer, 2017; Shaffer, Collier, & Ruis, 2016; Shaffer & Ruis, 2017) to data from *RescuShell* using the ENA Web Tool (version 1.5.2) (Marquart, Hinojosa, Swiecki, Eagan, & Shaffer, 2018).

The ENA technique has been described in detail elsewhere (Shaffer, 2017; Shaffer, Collier, & Ruis, 2016), including in previous papers presented at CSCL and ICLS (Collier, Ruis, & Shaffer, 2016; Csanadi et al., 2017; Siebert-Evenstone et al. 2016; Swiecki, & Shaffer 2018). Briefly, ENA models connections in discourse as a network graph where nodes correspond to the codes, and edges reflect the relative frequency of co-occurrence, or connection, between two codes. These networks can be further compared both quantitatively and qualitatively to analyze and compare different discourse patterns in CSCL environments.

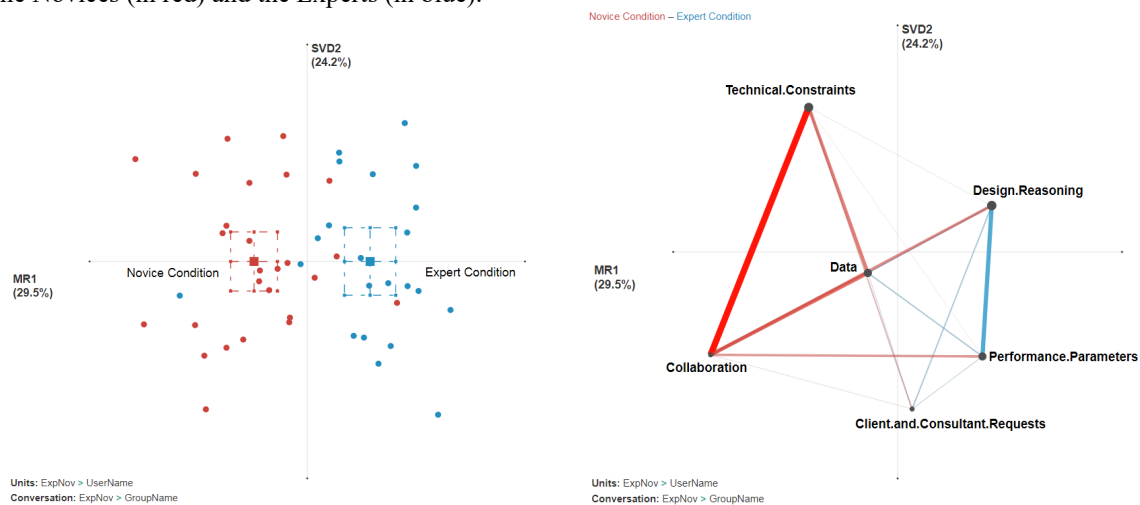
We are not providing the details of the data and coding scheme here, both because they are available in Chesler et al. (2015) and because the specific codes and details of the model are not the primary issue in testing the sensitivity to of the model to coding error. However, briefly: for the Gold-Standard result, we defined the units of analysis as all lines of data associated each individual student, and used a moving window to construct a network model for each line in the data, showing how codes in give line are connected to codes that occur within

the recent temporal context (Siebert-Evenstone et al., 2017), defined as 7 lines (each line plus the 6 previous lines) within a given conversation. The resulting networks were aggregated for all lines for each unit of analysis in the model. In this model, we aggregated networks using a binary summation in which the networks for a given line reflect the presence or absence of the co-occurrence of each pair of codes within the recent temporal context of the line. Our model included the following codes: CLIENT AND CONSUMER REQUESTS, COLLABORATION, DATA, DESIGN REASONING, PERFORMANCE PARAMETERS, and TECHNICAL CONSTRAINTS. We defined conversations as all lines of data associated with a single group in a single activity in each condition, Experts or Novices. We projected the resulting networks using a *means rotation*: a dimensional reduction technique that places the greatest differences between conditions on the x-axis in the ENA space.

In the resulting network graphs, the nodes were placed in ENA space so as to minimize the sum of the distances for each student between the point representing the student’s network in the projected space and the centroid of the student’s network graph. The result is two coordinated representations for each student: (1) a plotted point, which represents the location of that student’s network in the low-dimensional projected space, and (2) a weighted network graph, which shows the pattern of connections the student made that led to his or her location in the ENA space. Because of this co-registration, the positions of the network graph nodes—and the connections they define—can be used to interpret the dimensions of the projected space and explain the positions of the plotted points in that space. It is also possible to interpret the differences between two groups by constructing their mean networks and subtracting them. The resulting *difference network* shows the connections that are stronger for each group compared to the other.

Our model had co-registration correlations of 0.99 (Pearson) and 0.99 (Spearman) for the first dimension and co-registration correlations of 0.99 (Pearson) and 0.99 (Spearman) for the second dimension. These measures indicate that there is a strong goodness of fit between the visualization and the original model, and thus indicates that the model has interpretive validity: we can use the network graphs to interpret the statistical results.

*Test result.* To test for discourse pattern differences between Novices and Experts, we applied a two-sample t-test assuming unequal variance to the mean location of points in the projected ENA space for each condition. Figure 1 shows the plotted points (dots), means (squares), and confidence intervals (dashed boxes) for the Novices (in red) and the Experts (in blue).



**Figure 1.** Plotted points (left) and difference network (right) for the Gold-Standard ENA result.

Figure 1 also shows the difference network that can be used to interpret statistical differences between conditions. Where the difference network is red, the Novices had stronger connections, and where the network is blue, Experts made stronger connections.

Along the x axis, a two sample t-test assuming unequal variance showed the discourse pattern of Novices (mean = -0.37, SD = 0.44, N = 27) was statistically significantly different at the alpha=0.05 level from those of Experts (mean = 0.36, SD = 0.49, N = 28;  $t(52.78) = -5.76$ ,  $p < 0.01$ , Cohen’s  $d = 1.55$ ) (see Figure 1).

This statistically significant result (Figure 1, left) can be interpreted using the subtracted network graph on the right side of Figure 1. The Novices made more connections between TECHNICAL CONSTRAINTS and COLLABORATION, TECHNICAL CONSTRAINTS and DATA, COLLABORATION and DATA and COLLABORATION and PERFORMANCE PARAMETERS when compared with the Experts. The Experts made more connections between

PERFORMANCE PARAMETERS, and DESIGN REASONING than Novices. While Experts did make connections to COLLABORATION, more of their discourse focused on the relationships between PERFORMANCE PARAMETERS, DATA, and DESIGN REASONING; in other words, they were more focused on the core elements of engineering design. This result aligns with previous expert-novice comparisons from engineering virtual internships (Arastoopour et. al. 2016; Chesler et al., 2015).

Our Gold-Standard model thus includes two components that we tested in this analysis: (a) the statistically significant difference between Experts and Novices in that model, and (b) the specific locations of individual students relative to the model.

### Monte Carlo approach

For our BRT analysis, we created 5,000 simulated datasets at each of the different levels of coding error we examined (5%–10%, inclusive, in 1% increments)<sup>(3)</sup>. To create each individual simulated dataset, at each level of coding error, we randomly introduced error at that level to the Gold-Standard dataset. All other aspects of the subsequent ENA model creation process were the same<sup>(4)</sup>.

In the original data, codes are represented as a 1 or 0, indicating the presence or absence of a particular construct, for each turn of talk as students collaborated in groups of four<sup>(5)</sup>. We introduced coding error as follows: (a) for each code, we randomly selected a percentage of lines equal to the percent error rate; (b) we changed the coding of each line chosen, changing its value either from 1 to 0 or 0 to 1; and (c) we repeated this process for each code in the dataset.

### Sensitivity analyses

*RQ1: How much coding error can be introduced to a dataset before more than 5% of the 5,000 ENA simulated models do not find a statistically significant difference between the Novice and Expert conditions?*

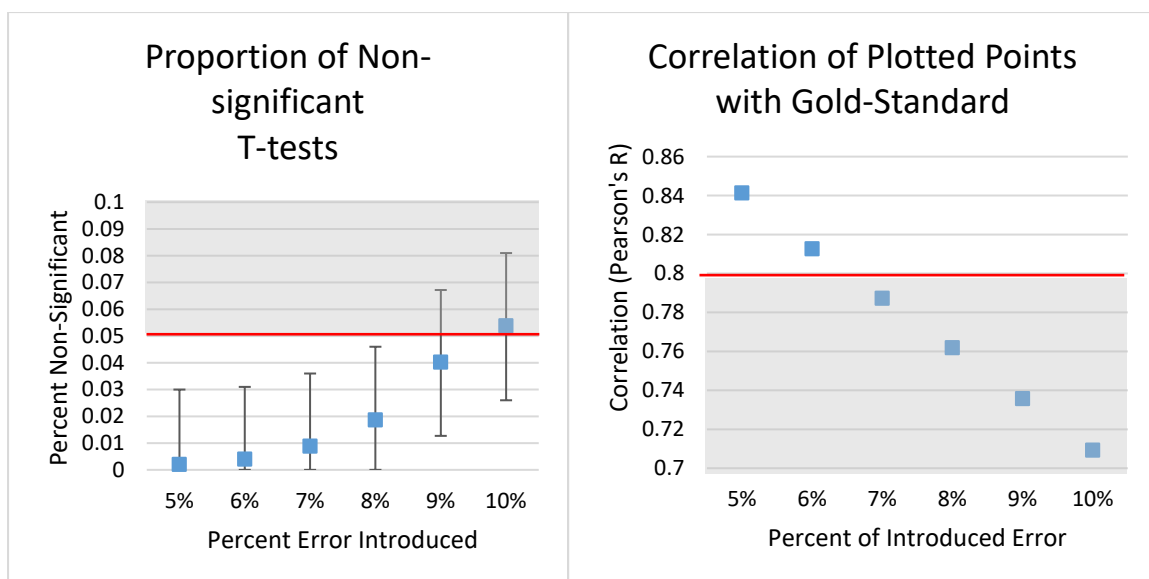
To answer research question 1, at each level of error: (1) we produced an ENA model for each of the 5,000 simulated datasets; (2) for each of these models we conducted the same t-test used to analyze our Gold-Standard model; (3) we calculated the number of models where the t-test *did not* show a statistically significant difference between Novices and Experts; (4) we created a 95% confidence interval on the estimate of the number of models that were not statistically significant. This tested the null hypothesis at each level of error: for example, at 6% introduced error, the null hypothesis was  $H_0$ : The difference between Novices and Experts is not significant at 6% coding error. If the confidence interval for the number of non-significant tests was below 5%, we rejected the null hypothesis and concluded that, with  $\alpha = 0.05$ , the group difference in the Gold-Standard result was robust to that level of introduced error.

*RQ2: How much coding error can be introduced to a dataset before the average correlation between the plotted points from the 5,000 simulated ENA models and the plotted points from the Gold-Standard model falls below 0.80?*

To assess the extent to which coding error affects the accuracy of assessing individual students, at each level of error: (1) we calculated the correlation between plotted points in the Gold-Standard model—which represent the results of the ENA analysis for each student—and the plotted points in each simulated ENA model; (2) we computed the average correlation between plotted points in the Gold Standard model and each simulated dataset; (3) we created 95% confidence intervals for the average correlations. This tested the null hypothesis at each level of error: for example, at 6% introduced error, the null hypothesis was  $H_0$ : The positions of students in is not different from Gold-Standard model with 6% coding error. There is no universally agreed upon threshold for significant correlations; some researchers propose 0.70–0.90 as a highly positive correlation (Hinkle, Wiersma, & Jurs, 2003), while others consider correlations of 0.80–1.00 very strong (Evans, 1996). We choose a conservative threshold of correlations greater than or equal to 0.80. If the lower bound of the 95% confidence interval was greater than 0.80, we rejected the null hypothesis and concluded that, with  $\alpha = 0.05$ , the Gold-Standard result preserved the positions of students in the model at that level of introduced error.

### Results

The left side of Figure 2 shows the percentage of tests in the sensitivity analysis that were not statistically significant at each level of introduced coding error we tested. The upper bound of the confidence intervals is below 0.05 at error rates up to 8%; however the upper bound of the confidence interval is above 0.05 at a 9% error rate. Thus the group difference in the Gold-Standard result is robust up to an 8% rate of coding error.



**Figure 2.** BRT results group difference sensitivity (left), individual level accuracy (right). For each graph, if the 95% confidence interval is at all contained by the shaded area, we fail to reject the corresponding null hypothesis.

The right side of Figure 2 shows the average correlation between the plotted points of the models with introduced error and the plotted points from the Gold-Standard model at the different levels of coding error. Because of the large number of simulated data sets (5000) at each level of error, the 95% confidence intervals are too small to be seen on the graph. The upper and lower bound of the confidence interval for error rates up to 6% are above 0.80; however, at a 7% error rate 7%, the confidence interval falls below 0.80. Thus, accuracy of individual measurements of the Gold-Standard model is robust to 6% error.

## Discussion

Our results thus show that difference between the means of two groups of students in the result we tested was robust to 8% error, but the accuracy of the model for individual students was only robust to 6% error. We thus conclude that the BRT suggests the result overall was robust up to 6% error.

While this analysis was conducted with only one set of CSCL data and using only one CSCL modeling approach, we believe that our analysis provides an example of how the BRT method for conducting sensitivity analyses can be used in CSCL research. In particular, we argue that the BRT approach can be used to conduct a sensitivity analysis on any model that depends on binary-coded data. That is, BRT provides a method that could be used to measure how robust CSCL results are to coding error.

Measuring the sensitivity of CSCL results to coding error in this way would provide a useful tool for establishing confidence in CSCL models. For example, if there are two conflicting CSCL results and the first was robust up to 1% coding error and the other was robust up to 8% coding error, we might have more confidence in the more robust result. More important, however, the BRT approach potentially provides a method for determining what level of coding reliability would be required to draw conclusions from a model—that is, a way to determine empirically the appropriate IRR threshold for a given result.

There is, of course, more work that will need to be done to make it possible to use BRT in this way. In particular, this study has several clear limitations. First, while this study reports on the robustness of (a) discriminating between two groups and (b) the accuracy of individual assessment, our analysis did not investigate how coding error affected the *interpretation* of the model. While it is beyond the scope of this paper, we have conducted BRT analyses for the interpretation of this model. Briefly, ENA models are interpreted using the positions of the nodes in the network graphs. In a future paper we will show that the average correlation between ENA node positions from the Gold-Standard model and the node positions from each simulated ENA model was above our threshold of 0.80 up to a 10% error rate. This suggests that that model interpretation was robust to 10% introduced error, which aligns with previous methodological analysis of ENA (Ruis et al. 2018).

A second limitation of this work is that we did not systematically account for properties of the codes, such as code frequency, which may impact sensitivity. Similarly, when researchers code data, errors may be systematic, rather than random. In future work we plan to account for these issues in the BRT method by

independently specifying rates of Type I and Type II coding errors. Third, we used only one model to explore the BRT method, and we deliberately chose a result with a large effect size. In future work we plan to test the BRT method using other results with a range of effect sizes, as well as other modeling techniques used in CSCL research. Finally, we used percent agreement as our IRR metric, which is a problematic measure (Cohen, 1960; Shaffer, 2017). In future work, we plan to base our error rates on more precise measures of IRR, including Cohen's kappa, precision, recall, and F statistics.

Despite these clear limitations, the results presented here suggest that the BRT method is a promising method for empirically deriving IRR thresholds.

## Endnotes

- (1) In statistics, sensitivity analyses often focus on bias, or systematic error, but for the purposes of this study we are focusing on the effects of random error. If a method is shown to be negatively impacted by random error, it will most likely be even more negatively impacted by systematic bias. It follows that focusing on random error is a more conservative approach for this type of sensitivity analysis.
- (2) Note that this approach does not require that an actual ground truth exists; only that we can say something about likely rates of disagreement that exist about a coding process that tries to locate it. This is an important conceptual issue, but beyond the scope of the current paper.
- (3) We also created simulated datasets at coding error levels from 1% - 5% and above 10% but in this study we are only reporting where the sensitivity analysis demonstrated a change statistically significant impact of coding error.
- (4) We projected each ENA model into our Gold-Standard ENA space by using the same Eigen vectors for the ENA rotation matrix in the dimensional reduction step of the ENA process. This resulted in the node positions for all of the ENA models in the sensitivity analysis to be identical. The only thing that varied was the coding and resulting connection strengths and plotted point locations for each unit in each ENA space. In other words, the creation of each of the ENA sets based on datasets containing introduced coding error was identical.
- (5) ENA can handle non-binary or weighted data, but this study focused on binary ENA models.

## References

- Arastoopour, G., Shaffer, D.W., Swiecki, Z., Ruis, A.R., & Chesler, N.C. (2016). Teaching and assessing engineering design thinking with virtual internships and epistemic network analysis. *International Journal of Engineering Education*, 32(3B), 1492–1501.
- Andrews, T. M., Leonard, M. J., Colgrove, C. A., & Kalinowski, S. T. (2011). Active learning not associated with student learning in a random sample of college biology courses. *CBE—Life Sciences Education*, 10(4), 394-405.
- Chi, M. T. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *The journal of the learning sciences*, 6(3), 271-315.
- Chesler, N. C., Ruis, A. R., Collier, W., Swiecki, Z., Arastoopour, G., & Shaffer, D. W. (2015). A novel paradigm for engineering education: Virtual internships with individualized mentoring and assessment of engineering thinking. *Journal of biomechanical engineering*, 137(2), 024701.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20(1): 37–46.
- Collier, W., Ruis, A.R., & Shaffer, D.W. (2016). Local versus global connection making in discourse. *Paper presented at the 12th International Conference of the Learning Sciences*. Singapore.
- Csanadi, A., Eagan, B., Shaffer, D. W., Kollar, I., & Fischer, F. (2017). Collaborative and individual scientific reasoning of pre-service teachers: New insights through epistemic network analysis (ENA). In B. K. Smith, M. Borge, E. Mercier, & K. Y. Lim (Eds.), *Making a difference: Prioritizing equity and access in CSCL: 12th International Conference on Computer-Supported Collaborative Learning* (Vol. I, pp. 215–222).
- Eagan, B., Rogers, B., Serlin, R., Ruis, A. R., Arastoopour Irgens, G., Shaffer, D. W. (2017). “Can We Rely on Reliability? Testing the Assumptions of Inter-Rater Reliability,” *Making a Difference: Prioritizing Equity and Access in CSCL: 12th International Conference on Computer-Supported Collaborative Learning*, eds. B. K. Smith, M. Borge, E. Mercier, & K. Y. Lim (2017), II:529–532.
- Evans, J. D. (1996). *Straightforward statistics for the behavioral sciences*. Brooks/Cole.
- Frank, K., & Min, K. S. (2007). 10. Indices of Robustness for Sample Representation. *Sociological Methodology*, 37(1), 349-392.
- Goh, N. S., Desai, S. R., Veeraghavan, S., Hansell, D. M., Copley, S. J., Maher, T. M., ... & Bozovic, G. (2008). Interstitial lung disease in systemic sclerosis: a simple staging system. *American journal of respiratory and critical care medicine*, 177(11), 1248-1254.

- Harwell, M. R. (1992). Summarizing Monte Carlo Results in Methodological Research. *Journal of Educational Statistics*, 17(4), 297–313.
- Hinkle DE, Wiersma W, Jurs SG. Applied Statistics for the Behavioral Sciences. 5th ed. Boston: Houghton Mifflin; 2003.
- Marquart, C. L., Hinojosa, C., Swiecki, Z., Eagan, B., & Shaffer, D. W. (2018). Epistemic Network Analysis (Version 1.5.2) [Software]. Available from <http://app.epistemicnetwork.org>
- Lasorsa, D. L., Lewis, S. C., & Holton, A. E. (2012). Normalizing Twitter: Journalism practice in an emerging communication space. *Journalism studies*, 13(1), 19-36.
- Ruis, A.R., Siebert-Evenstone, A.L., Pozen, R., Eagan, B., & Shaffer, D.W. (2018). A method for determining the extent of recent temporal context in analyses of complex, collaborative thinking. In Kay, J. & Luckin, R (Eds.) *Rethinking Learning in the Digital Age: Making the Learning Sciences Count, 13th International Conference of the Learning Sciences (ICLS), III*, (pp. 1625–1626).
- Schwarm, S. E., & Ostendorf, M. (2005, June). Reading level assessment using support vector machines and statistical language models. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics* (pp. 523-530). Association for Computational Linguistics.
- Shaffer, D. W. (2017). Quantitative ethnography. Madison, WI: Cathcart Press.
- Shaffer, D. W., Collier, W., & Ruis, A. R. (2016). A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics*, 3(3), 9–45.
- Siebert-Evenstone, A.L., Arastoopour Irgens, G., Collier, W., Swiecki, Z., Ruis, A.R., & Shaffer, D.W. (2017). In search of conversational grain size: Modelling semantic structure using moving stanza windows. *Journal of Learning Analytics*, 4(3), 123–139.
- Swiecki, Z. & Shaffer, D.W. (2018). Toward a taxonomy of team performance visualization tools. In Kay J. & Luckin, R (Eds.) *Rethinking Learning in the Digital Age: Making the Learning Sciences Count, 13th International Conference of the Learning Sciences (ICLS), III*, 144–151.
- Viera, A.J., & Garrett, J.M. (2005). Understanding interobserver agreement: The kappa statistic. *Family Medicine*, 37(5), 360-363.

## Acknowledgements

This work was funded in part by the National Science Foundation (DRL-1661036, DRL-1713110), the Wisconsin Alumni Research Foundation, and the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison. The opinions, findings, and conclusions do not reflect the views of the funding agencies, cooperating institutions, or other individuals.

# Developing Relationships, Changing Participation: Computational Identity

Melissa Gresalfi, Vanderbilt University, melissa.gresalfi@vanderbilt.edu

Amanda Bell, Vanderbilt University, amanda.m.bell@vanderbilt.edu

Alexandria T. Cervantes, California University Monterey Bay, alecervantes@csumb.edu

**Abstract:** This paper examines a single case of an African American boy, Eric, who enrolled in a free week-long summer camp called *Art and Coding*. We consider how his participatory identity developed in relation to the design of tasks, the students and teachers with whom he interacted, and the tools that he leveraged. With the goal of better understanding the relationship between participation and the development of beliefs about oneself, we then connect our analyses with the students' self-reflections in interviews. For Eric, a novice programmer whose passion is art, the experience of being positioned routinely as a knowledgeable explorer supported him to feel capable of resolving problems and attaining goals that he set for himself, even as he encountered challenges and frustration.

## Introduction

As Computer Science (CS) becomes increasingly prevalent in K-12 contexts, and given what we know about participation with CS in higher education and beyond, there is a need to consider not only what students have opportunities to learn about CS, but also what kinds of identities students are developing in relation to the domain. As the subject of significant attention in STEM domains, identities are seen as inextricable from learning, decision-making, and persistence. However, because CS has, until recently, primarily been a discipline reserved for higher education, we know little about the ways that students develop relationships with the domain across the age span. We do know that the proportion of women in computing jobs has actually decreased since 1990, down to 25% in 2015 (NCWIT, 2017), and just 7% of workers in computing in 2014 identified as Black and 7% as Hispanic (Beckhusen, 2016). These patterns mirror those seen in Mathematics, a discipline to which much time has been devoted to better understand how to change persistent patterns of course taking, major selection, and career choice. Drawing from research in mathematics education, we can predict that CS, when introduced to students before Advanced Placement courses, could also have wildly disparate influences on learning, enjoyment, persistence, and identity, depending on how it is taught. Using the assumptions and frames developed in mathematics education as our starting point, we explore how situated identities can develop in relation to programming, one skill in CS. In this paper, we consider a single case of an African American boy, Eric, who enrolled in a free week-long summer camp called *Art and Coding*. We consider how his participatory identity developed in relation to the design of tasks, the students and teachers with whom he interacted, and the tools he leveraged. With the goal of better understanding the relationship between participation and the development of beliefs about oneself, we then connect our analyses with the students' self reflections in interviews.

## Theoretical framework

Identity has been conceptualized in many ways, some of which highlight narratives and beliefs (Larnell, 2016; Leyva, 2016; Sfard & Prusak, 2005), and some that highlight activity and behavior (Hand & Gresalfi, 2015; Nasir, 2002). There are good reasons to call both of these things "identity," as the use of the same word helps to emphasize the interconnectedness in what we do, how others recognize what we do, and the stories we tell ourselves and others about ourselves. However, there is also utility in making distinctions between behavior and beliefs, particularly when it comes to making sense of the role that contexts play in shaping both. In this work we emphasize questions about how participation shifts over time, and then connect this to the ways that students talk about or make sense of those shifts.

Focusing on participatory identities as they are constructed in settings can contribute to our emergent understanding of the strategies and approaches that are useful in teaching students about computer science and computational thinking. To do so we draw on conceptions of situated identities that highlight how people participate with a set of practices (Hand & Gresalfi, 2015; Nasir, 2002). Early work on identity from a situative perspective sought to document students' participation in relation to a set of practices, influenced by Lave & Wenger's (1991) claim that identity and learning are intertwined. Analyses of identity therefore require documenting how students are participating in relation to a set of classroom practices, and how those practices serve to create opportunities for identity development. These analyses are tightly coupled with studies of

classroom norms and practices and focus on interaction, with students as the unit of analysis (Grant, Crompton, & Ford, 2015; Langer-Osuna, 2015; Sengupta-Irving & Enyedy, 2015).

We draw on this framing of identity with the goal of better understanding how student identities are co-produced through their experience with the domain. We see students' decisions not to continue with a subject or not to pursue a major as co-produced through their experiences of the domain, and the resulting ways they view themselves in relation to the domain. Too often, differences in course taking are conceptualized as failures of confidence, of preparation, or of imagination. But research on identity suggests that continuing to persist with activities that are damaging come at great significant cost, and students are wise to bow out. Thus, a better understanding of how activities produce this damage—and how we might be able to change it—is essential to increasing participation in the domain in the long term.

## Designing to support identities in computer science

Although little research has focused on participatory identity in relation to CS, research from K-12 contexts have documented several design characteristics of CS learning environments that support the development of productive elements of identity, including communities, mentors and role models, collaborative work, and programming contexts like stories or games. First, research suggests that representation matters, both in terms of the content or models in activities, as well as through the images and objects in the room. For example, in *Digital Youth Divas*, researchers designed characters to imitate actual middle school girls with a variety of interests, body types, and stories (Pinkard, Erete, Martin, & McKinney de Royston, 2017). The *Digital Youth Divas* program also demonstrated the value of role models, whether real or fictional, for interest and personal identity construction. The relatable characters and situations offer ideational resources to support girls' identification with CS. Likewise, even reducing stereotypical objects in computing classrooms (e.g. replacing Star Wars posters, electronic parts, and tech magazines with art, plants, and general magazines) can increase girls' sense of belonging and interest in a high school computing course without lowering boys' existing interests (Master, Cheryan, & Meltzoff, 2016).

It is not only *who* students see however, but also *what they can do* that seems to support productive engagement and identity development. Highlighting the social nature of computing in environments like Scratch, with its online community, can create positive, gender-inclusive educational experiences for newcomers (Resnick et al., 2009). Likewise, teaching CS through programming stories, dances, and games can connect to students' interests and create gender-inclusive learning environments (Daily et al., 2014) that can increase the time spent persisting on programming projects (Kelleher, 2008).

## Context of the study

Data for this paper comes from a free five-day summer camp for middle school students, held in a southeastern U.S. city. The camp was titled "Code Your Art," and was advertised as involving computer programming and art. There were two classrooms during the camp, with 16 students in each class. Each class was co-caught by two teachers. The teachers taught elementary or middle school mathematics during the regular school year. There were also six researchers and two teaching assistants present during the camp, rotating among the classrooms as needed.

The goal of the camp was two-fold. First, the project was funded to explore how mathematical thinking might co-develop or connect with the algorithmic thinking that is central to programming. Second, we wanted to understand how the design of tasks and activities invite participation from students who might not see themselves as programmers or have any prior experience with programming. To meet both needs, we designed a set of activities that were sufficiently open to be seen by students as spaces for design and self-expression, but sufficiently constrained to make algorithmic thinking useful. Image design, with its attunement to position and movement, was an ideal fit for that goal. Complimenting the focus on design was the appeal of deliberately naming and emphasizing art in the camp as a signal to a broad range of students that this camp was intended for them, and mindful of the ways programming, building, and engineering often signal masculinity (Butler, 2002; Master et al., 2016).

The five days ultimately included a variety of activities involving computers, physical materials, and embodied activities. The programming activities used NetLogo, a multi-agent environment developed by Uri Wilensky and his colleagues (Wilensky, 1999). We chose NetLogo because it had functionality to import and edit images and included multiple, accessible ways to think about modifying images (thinking about patches and turtles, the latter offering opportunities for dynamic animation). Because NetLogo uses text-based language for programming, we developed a set of introductory models that included buttons that allowed students to explore possible features and functions. Such buttons can be opened so that the code is shared, meaning that students can both explore what a button does and how it does it. During the camp students edited models, debugged existing models, and created projects from empty models, and then shared their work to a digital gallery that was curated



on a camp website. They also wrote daily blog posts and shared them on the camp website. Other activities included building with Perler beads and acting out programs with their bodies.

## Methods

For this first analysis, we focus on one student from the camp. Of course, this single case does not serve as a model for what happened to all students, nor is that the purpose of the analysis. Single case studies can be helpful when trying to understand a phenomenon in some depth; here, we seek to better understand how the activities of the camp served to position the student relative to content and to others. The focal student, Eric, was selected for further analysis because although he appeared to be engaged throughout the entire camp, he had no prior experience with computer programming, and he signed up for the camp because of his love of art. Eric sat in the same seat all days of the camp, primarily interacting with only one boy, Kyle. He tended to be quiet and focused, but was quick to share his thinking when the class was discussing ideas associated with art and artists. Eric was a rising 7<sup>th</sup> grader at the school where the camp was held, and he expressed that he enjoyed the camp and attended the camp all five days.

Multiple forms of data were collected during the camp; for this paper we rely primarily on screen captures of the students' programming, observation notes that were taken during the camp, and interviews that were conducted one-on-one at the beginning and end of the camp. Interviews were semi-structured, and the protocol focused on initial reactions to and ideas about the camp, their reflection about their experiences during the camp, and how they see themselves and others within computer science. To analyze these data, we first reviewed the screen capture software taken of Eric, with the goal of attempting to understand his approach to programming. We categorized his work by attending to the ways he was positioned relative to the activity, for example by noting how he explored the software, how he refined and changed his designs as they developed, and whether he persisted in the face of challenge. We also looked for the ways that Eric was positioned relative to others, for example whether he asked for help, was asked for help by peers, or made comments about himself or others that indicated some sense of relative ability or skill level.

## Findings

In what follows, we describe characteristics of Eric's shift in participation with the tools and resources over time, and also how his thinking about himself (or more typically, activities) changed over time.

### Positioning with respect to the activities: From exploration to intention

From the very first day of camp, Eric spent a long time exploring and modifying the NetLogo models he was given, and experimenting with their functionality in relation to the images he chose to use for his projects, which primarily involved logos and images of basketball. However, his exploration at the beginning of the camp was less obviously guided by the kinds of disciplinary understandings that characterized his later work.

#### Day 1

In one of the first models students saw, they were asked to select an image from a wide collection created by the research team, and consider how they might want to modify it in a way that was interesting to them. The instructions were intentionally open with respect to what interesting might mean, or how the image might be changed. The initial model included multiple buttons that operated on an image using color (changing all colors, changing some colors, changing colors randomly, etc.) These buttons had different functions but could also relate to one another. In creating buttons that were interactive, we sought to offer a resource for students to inquire into how the models worked in general, and how modifications to images could be made that built on one another.

As an example of how Eric's exploration changed, we examined his use of a *variable slider* on Day 1 and Day 4. Variable sliders are a type of interactive button that allows the user to change the value of a stored variable. The variable is referenced in other buttons in the model. The user can adjust the slider then click the button that references the variable, repeating the process to see how changing the variable changes the output. The models Eric used on days one and four both had variable sliders, but they referred to different variables. On day one, the teacher first introduced the model to the class and showed how to adjust the canvas size, and she explained that they should explore and edit an image. Eric then played with the model for 30 minutes before the below episode occurred.

We characterize Eric's initial use of the model as open (as opposed to systematic) exploration, as there is a lack of connection between the ways that he used buttons on the model. For example, in Figure 1 cell A, Eric clicked the button *ask patches with [pcolor = black] [set pcolor color#]*. At the time, the slider that read *color#* was set to 65. He might have expected the part of the canvas that appears to be black to change to a different color, but he clicked the button three times and nothing changed. This is probably because the patches were not

exactly black (as read by NetLogo). Eric's next move (cell B) suggests that he suspected the lack of change was due to the value in the slider for the *color#* variable, as he slid the variable to 115. Moving the slider had no effect, as the slider changed the variable but (as is designed to do), did not apply it to the canvas. Following this move, we might expect that Eric would then click a button that references *color#*, as in step 1, to see how the variable change affects the image. Instead, Eric clicked a button that did not use the *color#* variable. This button, *ask patches [set pcolor (140 - pcolor)]* changes the color of all patches. He clicked the button twice, so first the canvas changed to blue and red (cell C), then returned back to the original colors.

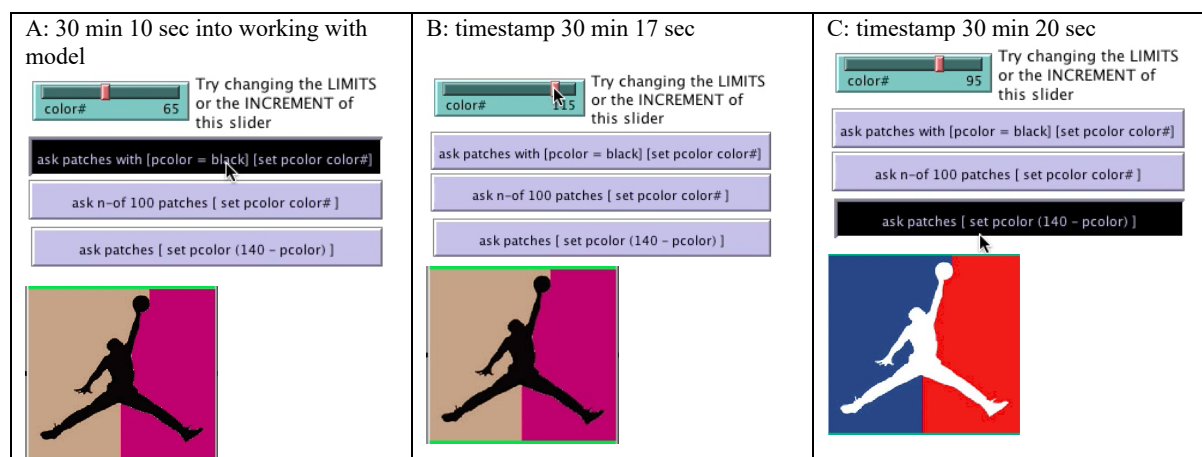


Figure 1. Three sequential buttons clicks of a model on day 1.

This sequence of button clicking suggests that Eric either did not notice, or did not appreciate the significance of the connection between the label on the variable (*color#*) with buttons that reference that variable. Indeed, subsequent moves confirm that Eric's exploration did not seem to be attuned to this coupling. In cell D in Figure 2, Eric adjusted the *color#* variable slider a second time. Again, we would expect him to click a button that uses *color#* after this to see how changing the variable affects the output. Instead, (cell E) Eric clicked the same button again that did not use the variable: *ask patches [set pcolor (140 - pcolor)]*. Again, he clicked the button twice so it cycled the canvas back to the original colors. Since he repeated the same actions and nothing changed, it's not clear if he understood what the button was doing. Finally, Eric clicked another button that referenced *color#*: *ask n-of 100 patches [set pcolor color#]* and repeatedly clicks this button over 20 times (cell F). He might have noticed what the button was doing (changing random patches to a different color), but it is not clear if he understood how it connected to the *color#* variable.

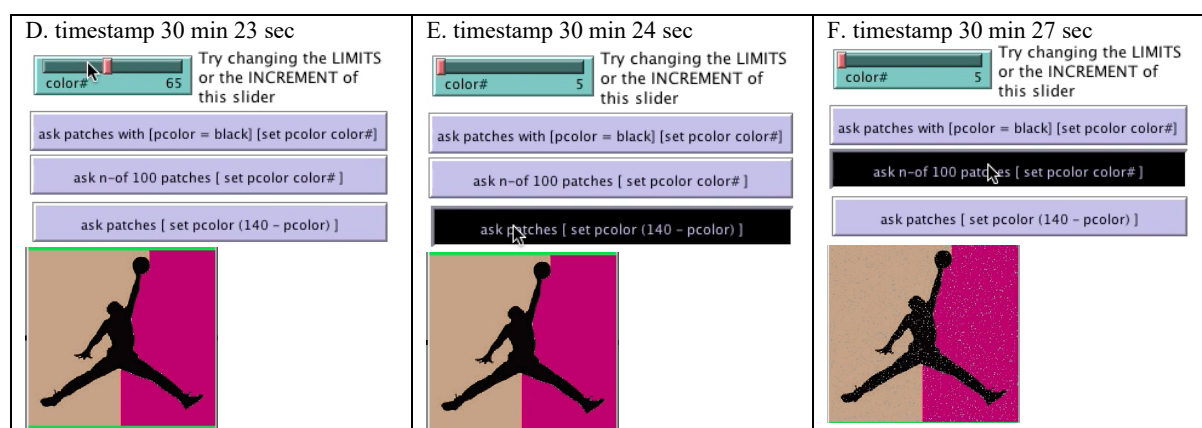


Figure 2. Three subsequent button clicks of a model on day 1.

Eric explored this model for almost 90 minutes, indicating his interest in, and commitment to, the activity. However, the way he explored the model suggests that he was not guided by the notion of variable, as he did not seem to yet have access to the appropriate disciplinary knowledge that would lead him to see connections between different buttons. This was just a short example, but during the whole hour and a half session with NetLogo on day one, we did not see systematicity develop in his approach to the model. However, it is important to note that

Eric stayed engaged with the model, repeatedly choosing basketball-related images and exploring how the image changed in relation to the buttons, suggesting that he was enjoying the activity, engaging in work that was in some way satisfying to him, and developing a relationship to the activity that emphasized his own agency and exploration.

#### Day 4

Over the course of the week, Eric began to engage and explore models more systematically, even when the concepts in the model were entirely new. For example, on day 4 a researcher introduced several new models, to offer a vision of what some “cool models” could do that the students might find useful or inspiring for their final projects. As on day one, Eric had not used this model himself previously. However, Eric’s exploration of the model seemed more noticeably strategic. After opening the model for the first time, Eric imported an image, and then saved the canvas to a variable by clicking *save current colors in patch memory mem1* (Figure 3, cell A). In doing so, he makes use of something that he has learned in previous days. Eric then imported a second image (cell B) and saved the second canvas to another variable by clicking the *save current colors in patch memory mem2* (cell C). He used the two save buttons appropriately so that each image was saved in a separate variable. These moves suggest that he has learned how these two variables work, as he is leveraging concepts he has used on previous days. However, his next moves indicate his more intentional exploration of the model.

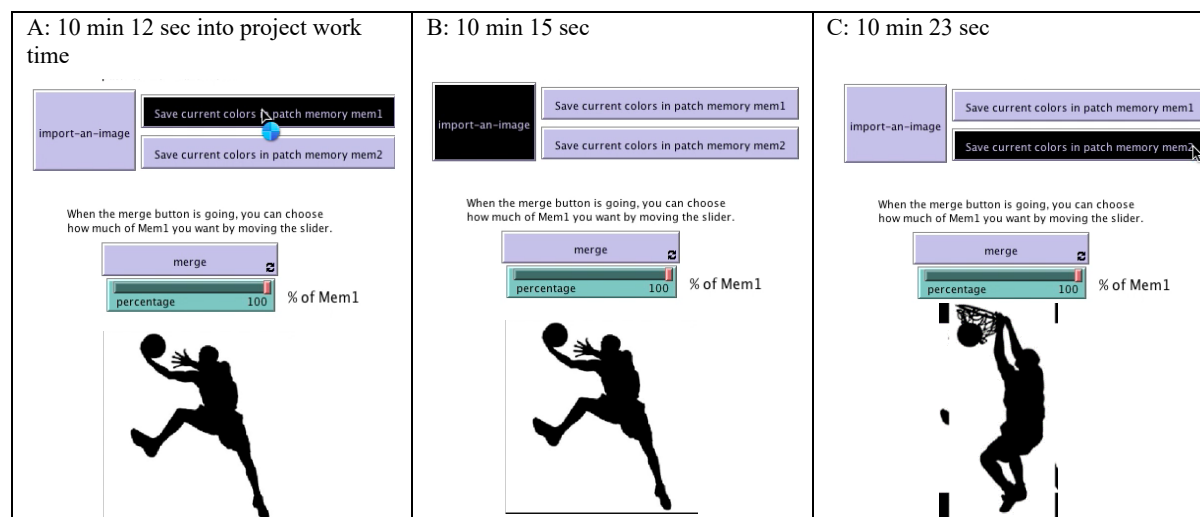


Figure 3. Four sequential buttons clicks of a model on day 4.

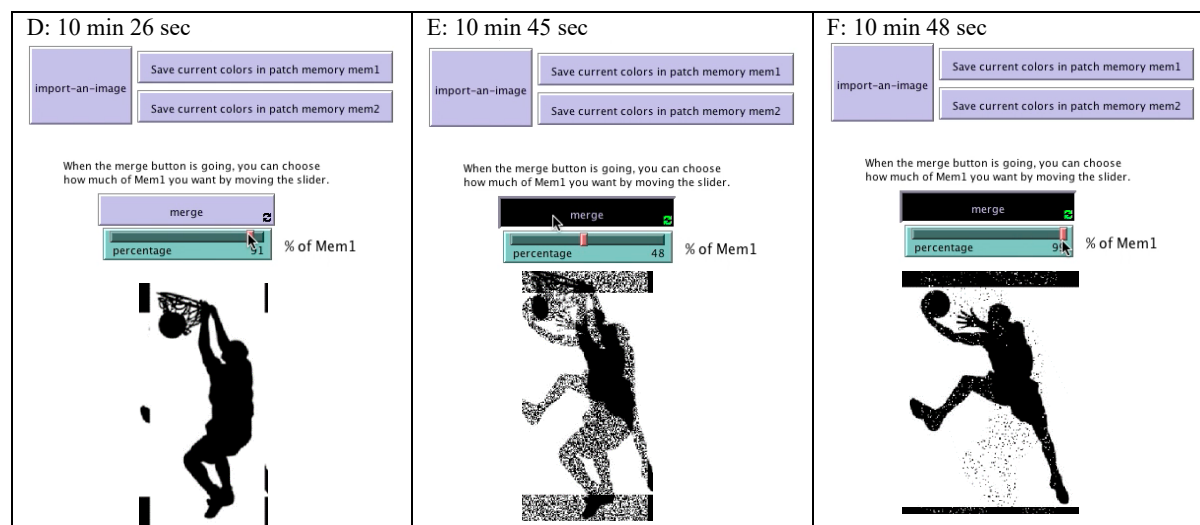


Figure 4. Eric’s exploration of a model on Day 4.

In Figure 4, Eric adjusted the *percentage* variable slider (cell D). After noticing no change, Eric paused for 19 seconds, looking intently at the model, then said “Oh I have to merge it.” He clicked the *merge* button (cell E), suggesting that he deduced that the “merge” button used the *percentage* variable even though it was not written in the name of the button. With the merge button on, Eric adjusted the slider again and noticed the images fade from one to the other (cell F). Eric said “yeah that’s really cool. Alright I can do that.”

This episode suggests a shift in the way Eric approached this model, supported by both a change in his own understanding of the models and the affordances of the programming environment. The strongest characterization of his participation is that of intentionality. This can be seen in the model he chose to investigate, the images he selected, and his exploration of the model itself. Eric selected an image blurring model, and then chose the images he imported with a purpose—the images in cells 2 and 3, when blended, create an animation that looks like a player dunking the basketball. Eric found these images online himself. His use of the model indicates a shift in his understanding of variables: Eric was methodical as he imported an image, saved it, then imported another image and saved it. This systematicity might seem simple and straightforward, but amongst camp participants, understanding what “patch memory” was, and how to use that feature appropriately, was often confusing and required multiple iterations. Next, Eric adjusted the slider variable. When he used the variable slider on day 1, he subsequently clicked several buttons that did not reference the variable, and he repeated that process multiple times. It appeared that Eric was exploring without really understanding what the buttons were doing or how they related to one another. On this day, Eric paused after he adjusted the slider and noticed no change. He looked around the screen, appearing to take the time to read the other buttons to see what he needed to do next. After pausing, he clicked the “merge” button. With the merge button pressed, he adjusted the slider again and noticed the image change. It appeared that he noticed the connection between the variable slider and the merge button in this model, unlike the way he used the slider on day 1. His exploration of this model was much more methodical and goal-oriented than on day 1, since he imported particular images and had an idea of what he wanted to happen, so Eric just had to figure out how to reach his goals with the buttons he was given. On day 1, he imported an image and spent the rest of the time clicking around different buttons to see what they did, but he did not appear to draw connections between any of the buttons given the order in which he clicked on different things throughout the session.

### Positioning with respect to others: From keeping quiet to mobilizing assistance

A second resource that supported and demonstrated Eric’s change in participation could be seen in the ways he marshalled different forms of assistance. In day one, Eric seemed to attempt to be as inconspicuous as possible: he was generally quiet and talked very little, and did not speak to an adult until directly approached and being asked if he needed help. He replied by saying he was alright or “just playing.” This kind of exchange happened between him and adults in the camp several times early on. Below is an example of one of those instances; Eric was focused on his computer, looking at a NetLogo model:

Teacher: *Do you need some buttons, Eric?*

Eric: *I was just ah, I don’t know.* (Eric navigates to the camp website)

Teacher: *Just playing?*

Eric: *Yeah. I’m trying to figure out where that, okay.* (Eric re-opens his NetLogo file)

(the Teacher stands behind him for a few seconds, then walks away)

In contrast, Eric made strategic use of adult expertise on the fourth day. Below, Eric appeared to know what he wanted to accomplish with the model, and he asked for help when he was stuck along the design process:

Eric: (talking to himself) *That’s really cool. I’m still gonna need help, though.* (calling out to the room) *Could you help me please?* (a researcher comes over) *So I want to add more than one image here. How would I do that?*

Researcher: *Yeah! So...* (continues to explain how to add a third image; Eric does the clicking or typing on the computer while the researcher gives instructions)

Overall, over the course of the week we see Eric’s participatory identity evolving, from engaged compliance to intentional design. This shift appeared to be supported by a set of tasks that were open and positioned students to be the ones to determine what they wanted to make, what “story” they wanted to tell, or what effect they were trying to create. For Eric, such openness allowed him to explore and use images that

appeared to be personally meaningful, but that also could be used in ways to create visually pleasing effects. Over the week, Eric demonstrated more confidence and intention, thinking strategically when trying to understand how something worked, and gaining significant personal insight into what he knew how to do and what he would need help accomplishing. It seemed that, for Eric, having the space to explore independently, and ask for help when needed, was supportive and motivating for him, as he described in his interviews.

## Reflections on activity

Eric's interviews were remarkably consistent with our observations of his participation in the camp. In his initial interview, despite claiming to know little about computer science or any computer scientists, Eric routinely described computer science as involving creativity, art, and expression. He assumed that computer scientists were confident, creative people. He claimed that he signed up for the camp because of his love of art, but did not seem overly concerned about his ability to engage with programming. Thus, although he was a novice, he did not appear to consider himself at any particular disadvantage or unlikely to succeed. This was consistent with his enthusiastic commitment that we observed early on in the camp, when he would persist in exploring models for over an hour, even when there was significant evidence that he didn't yet appreciate what the models could do.

In his final interview, Eric talked at some length about the camp being enjoyable and fun, but more importantly, described his experiences in the camp that add nuance to our observations of his activity.

Int: Okay. What is one thing you made this week that you were proud of?

Eric: The moving dunk [*referring to the project he started on day 4 in figure 2 above*] that I made on the computer because it was very hard and whenever it was difficult, I didn't give up and I just kept asking for help and trying to explain how I'd be able to do something and I eventually got to do it.

Int: Nice. So, what is something that was really frustrating this week?

Eric: Making the dunk change, a camera flash to different angles. It was probably not probably, it was the most difficult.

Int: It was? And how did you overcome that challenge and frustration?

Eric: I kept asking, I kept trying different possibilities until I found the right one.

In this exchange, Eric describes the project he began in Figure 2, narrating his own agency in asking for help and figuring things out, which we also observed. When asked what he did to persist in the face of challenge, he described his exploration of NetLogo, which again describes, retrospectively, the same behavior we observed in the screen captures. Thus, Eric demonstrated a change in his participation with coding that reflected not only a deepening familiarity with some key concepts, but also a more agentic and empowered relationship to the activities, where being stuck or not yet knowing how to realize his vision was not perceived as a problem.

## Discussion and implications

This single case study offers a picture of how a productive identity might develop in relation to activities, norms, and supports. For Eric, a novice programmer whose passion is art, the theme of the camp was inviting and made him feel welcome and capable. As Eric articulated, the experience of being positioned as a knowledgeable explorer supported him to feel capable of resolving problems and attaining goals that he set for himself, even when doing so required significant time and effort. Notably, Eric was readily able to identify challenges that he faced and times of frustration, but these were not perceived as setbacks. As Eric explained, and demonstrated repeatedly throughout the week, getting frustrated was no problem as long as he felt that he knew he would be able to resolve it. Over the course of the week, his strategies grew increasingly sophisticated, and he grew increasingly comfortable asking for help when he needed it.

A single case study is never intended to demonstrate what happens generally, but rather is useful as a means of exploring the mechanisms that support a phenomenon of interest—here, participatory identity in relation to CS. This analysis suggests that being positioned as a problem solver from the beginning of the week supported Eric to develop a relationship with the activities that allowed for tinkering, iteration, and even failure in order for a final product to emerge. As reflected by Eric himself, these activities were “unbelievably fun” and worth pursuing in the future. This seems a promising start to a productive identity in relation to CS. This was, of course, a short intervention, and this analysis considers a single case. We would not claim that this single week-long camp is likely to change students' long-term relationship with the domain, just as we do not claim that Eric's experience in the camp was the same as every other student's experience.

However, there are many parallels to Eric's experience in the camp, and his reflection on those experiences, with findings from the mathematics education literature that helps us to see this single case as offering a *petite generalization* (Stake, 1995) for the field. Research on the relationships that students develop with mathematics has demonstrated repeatedly that when students have opportunities to explore, reflect, and discuss in contexts where competence it is seen as being able to make strong arguments as opposed to answering questions quickly, a wider range of students claim to enjoy mathematics and persist in the discipline (Boaler & Greeno, 2000; Nasir, 2002). This example thus offers a suggestion that designing for productive k-12 identity development in computer science, though a new challenge to our field, can look to a wealth of answers from other more established fields for clear suggestions of how to proceed—and what to avoid.

## References

- Boaler, J., & Greeno, J. G. (2000). Identity, agency, and knowing in mathematics worlds. *Multiple perspectives on mathematics teaching and learning*, 171-200.
- Butler, J. (2002). *Gender trouble*: Routledge.
- Daily, S. B., Leonard, A. E., Jörg, S., Babu, S., & Gundersen, K. (2014, March). Dancing Alice: exploring embodied pedagogical strategies for learning computational thinking. In Proceedings of the 45th ACM technical symposium on Computer science education (pp. 91-96). ACM.
- Grant, M. R., Crompton, H., & Ford, D. J. (2015). Black male students and the Algebra Project: Mathematics identity as participation. *Journal of Urban Mathematics Education*, 8(2).
- Hand, V., & Gresalfi, M. (2015). The joint accomplishment of identity. *Educational Psychologist*, 50(3), 190-203.
- Kelleher, C. (2008). Using storytelling to introduce girls to computer programming. *Beyond Barbie & Mortal Kombat: New perspectives on gender and gaming*, 247-264.
- Langer-Osuna, J. M. (2015). From getting “fired” to becoming a collaborator: A case of the coconstruction of identity and engagement in a project-based mathematics classroom. *Journal of the Learning Sciences*, 24(1), 53-92.
- Larnell, G. V. (2016). More Than Just Skill: Examining Mathematics Identities, Racialized Narratives, and Remediation Among Black Undergraduates. *Journal for Research in Mathematics Education*, 47(3), 233-269.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. New York: Cambridge University Press.
- Leyva, L. A. (2016). An intersectional analysis of Latin@ college women's counter-stories in mathematics. *Journal of Urban Mathematics Education*, 9(2).
- Master, A., Cheryan, S., & Meltzoff, A. N. (2016). Computing whether she belongs: Stereotypes undermine girls' interest and sense of belonging in computer science. *Journal of Educational Psychology*, 108(3), 424.
- Nasir, N. S. (2002). Identity, goals, and learning: Mathematics in cultural practice. *Mathematical Thinking and Learning*, 4, 213-248.
- Pinkard, N., Erete, S., Martin, C. K., & McKinney de Royston, M. (2017). Digital Youth Divas: Exploring narrative-driven curriculum to spark middle school girls' interest in computational activities. *Journal of the Learning Sciences*, 26(3), 477-516.
- Resnick, M., Maloney, J., Monroy-Hernández, A., Rusk, N., Eastmond, E., Brennan, K., . . . Silverman, B. (2009). Scratch: programming for all. *Communications of the ACM*, 52(11), 60-67.
- Sengupta-Irving, T., & Enyedy, N. (2015). Why engaging in mathematical practices may explain stronger outcomes in affect and engagement: Comparing student-driven with highly guided inquiry. *Journal of the Learning Sciences*, 24(4), 550-592.
- Sfard, A., & Prusak, A. (2005). Telling identities: In search of an analytic tool for investigating learning as a culturally shaped activity. *Educational Researcher*, 34(4), 14-22.
- Stake, R. E. (1995). *The art of case study research*. Sage.
- Wilensky, U. 1999. NetLogo. <http://ccl.northwestern.edu/netlogo/>. Center for Connected Learning and Computer-Based Modeling, Northwestern University. Evanston, IL.

# Family Collaboration in the Digital Age: Parent Learning Partner Roles Are Linked to Child Expertise and Parents' Work

Brigid Barron, Caitlin K. Martin, and Judy Nguyen  
barronbj@stanford.edu, ckmartin@stanford.edu, judynguyen@stanford.edu  
Stanford University

**Abstract:** Our quantitative study investigates how parents' use of technology in their work relates to their roles as learning partners for their children. We also examine whether parent learning partner roles are associated with children's opportunity to experiment with digital production activities. Based on earlier ethnographic studies we predicted that there would be variability of the number of roles played across our sample, and that parents active in the technology industry would play more roles, especially those potentially catalyzed by some level of expertise or knowledge, such as teaching, modeling practice, collaborating on technical projects, or lending resources. Our findings were consistent with these hypotheses. For both mothers and fathers, the level of technology use at work was associated with the diversity of learning partner roles. Counts of the diversity and density of parent roles was significantly associated with the breadth of children's experience with digital production activities. Implications for equity are discussed.

## Introduction

Jonathon, a 13-year-old student in Silicon Valley, developed the front and backend of websites including linked databases and ran his own computer consulting business. Much of his knowledge developed at home as he tinkered and taught himself new programming languages, using the Internet and books as resources and collaborating with his parents on projects. Books, computers, and software were at the ready, lent and purchased by his parents, and broad expertise was available in his family and their network of friends and colleagues to help him advance his goals. His father was an engineer and programmer who introduced him to educational computer games in preschool and continued to teach him about computers and programming until his son's knowledge surpassed his own in some areas. His mother was a small business owner and contributed her expertise to Jonathon's own money-making ventures. In another Silicon Valley neighborhood was 12-year-old Andres, an avid gamer who became an expert in game design and graphic tools that allowed him to create characters and backgrounds for his game environments. Andres accessed resources and learned most of what he knew at a Computer Clubhouse space in a local community center and brought that knowledge home to his family. At home, he taught his mother how to use the mouse and keyboard and organize documents within folders and helped her to figure out how to pay bills online.

These examples from qualitative work illustrate some of the ways networked technologies and digital tools provide rich opportunities for novel and varied forms of intergenerational learning with family members dynamically taking on roles as teachers, learners, collaborators, and brokers (Barron, Martin, Takeuchi, & Fithian, 2009). Sources of variability in family-based learning opportunities are particularly important to understand from an equity perspective. There is growing evidence that expertise with digital technologies is related to further networked learning such as job opportunities (Peng, 2017), the tendency to use technology to learn (Horrigan, 2016), and teaching friends and family to use technology (Barron, Pinkard, Gomez, & Martin, 2014). Recent national data suggests that opportunities for computational learning are unevenly distributed (Google Inc. & Gallup Inc., 2016). Communities with fewer technology learning opportunities and knowledgeable networks may need additional supports while those with parents in jobs that require the use of technology or that involve its design may be in a particularly powerful position to support their children's digital hobbies that will prepare them for empowered and critical use. In the examples above, Andres and Jonathon had similar levels of interest but vastly different access to resources at home and in school (Barron, 2010). In this paper we report on a study of how the supporting roles parents play in the lives of their children vary with parent use of technology in their occupations.

The 2019 CSCL conference theme of *combining embodied, enactive, extended, and embedded learning in collaborative settings* (4EC) highlights the complexity of cognitive phenomena, encouraging analysis and understanding at a comprehensive systems level. Family networks are important as they represent the place where children spend much of their time and parents and other adults at home engage in guided participation as informal teachers and resource lenders (Rogoff, 1998). How and when adults and children take up roles as more expert learning guides at home is influenced by their cultural repertoires of practice around teaching and learning

(Gutiérrez & Rogoff, 2003) as well as opportunities and needs that arise due to rapidly changing technologies. In this paper, we consider the family as a networked learning ecosystem. That is, family members are connected to each other and to their own broader environments through learning opportunities at work, school, or in the community and these influence interactions at home. We explore the interactions between people as *enactive* elements of the system and look at those interactions through the lens of *extended* environments, in this case parent jobs that influence family learning within a child's broader learning ecology. We close with implications for future research and family learning opportunities.

## Parents as learning partners

There is evidence that being technologically fluent is increasingly important. Studies suggest that youth engage in gaming and social networking more than creative digital production activities (Barron, 2006; Ito *et al.*, 2009) but that these production activities using technology to create are linked to important competencies. Research focused on middle school students found that greater involvement in technology production experiences was significantly related to more confidence in their own capacity to generate new ideas, stronger self-efficacy with professional tools, a propensity to share technical expertise with a broader network of people, and intent to continue to develop technical expertise in the future (Barron & Martin, 2016; Martin, Barron, Stringer, & Matthews, 2014).

There is extensive research on how children learn through family interactions in areas such as language, academic skills, and traditional crafts, and in most cases, parents pass on their more expert knowledge (Rogoff, 2003; Whiting & Whiting, 1970). We know less about how young people learn and build fluencies with these newer technologies and how families fit in. Recent large-scale survey research in the UK suggests that active engagement with media by youth, such as creative production, is linked to parents' higher confidence using technology while more restrictive mediation of technology at home is correlated with parents with less confidence (Livingstone *et al.*, 2017). An in-depth case study of 22 Hong Kong students found positive links between parental supportive mediation and technology competence and youth generative use of information technology (Yuen *et al.*, 2018).

An ethnographic study of highly engaged middle school students committed to digital hobbies that ranged from robotics to making music videos found that parents played eight unique roles in surfacing, supporting, and coordinating learning opportunities for young people: teacher, learner, collaborator, learning broker, employer, financier, non-technical consultant, and resource lender (Barron *et al.*, 2009). Despite the fact that all youth were from one affluent Silicon Valley school, there was substantial variability in the level of parent engagement as indexed by the number of roles played in the household and the frequency with which these roles were played (for example across parents and over time). Greater parent involvement was associated with higher levels of child expertise and early onset (prior to age five) of a child's engagement in new media production activity. A number of smaller qualitative projects have validated these findings, particularly emphasizing how the practice of adults brokering for young learners, both in the community (Ching, Santo, Hoadley, & Peppler, 2016) and at home (Louw, Barbuto, & Crowley 2016) can expand youth opportunities and social networks. To validate and extend these findings, we created a survey measure of parent learning partner roles to administer to a larger sample.

## Research questions

In the current study we use survey data to ask (1) whether parent learning partner roles are associated with children's opportunity to experiment with digital production activities and (2) how parents' use of technology in their work relates to their roles as learning partners. Based on earlier ethnographic findings we predicted that there would be variability of the number of parent roles played across our sample, and that parents active in the technology industry would play more roles, especially those roles that are potentially catalyzed by some level of expertise or knowledge, such as teaching their child, modeling practice, collaborating on technical projects, or lending resources.

## Methods

### Participant sample

We surveyed the entire eighth grade cohort of students from a public middle school in the Silicon Valley region of California (N=366 students). The school was high performing (over 90% of students met or exceeded standard in math and English language arts on state tests) and affluent (only 3% were eligible for subsidized lunch programs). Approximately half of the students surveyed were male (51%). The sample had a high Asian-American population: 34% self-selected Chinese as their ethnicity, 26% Asian Indian, 23% White, 19% other Asian



(Japanese, Vietnamese, Korean), 5% Hispanic-Latino, and less than 1% African American (13% selected more than one option). Almost a quarter (23%) were born outside the United States, and 34% lived in multilingual households. Although the school was located within a hub of technological innovation, it did not offer computer science classes.

## Measures

Students were administered a survey of Access, Experience, and Interests in Computing (Barron, 2004; Barron, Walters, Martin, & Schatz, 2010) during their 40-minute mathematics period. In this paper, we report on a subset of measures including youth digital production experience, parent jobs, and our new parent learning partner roles item.

### Student digital production experiences

Students were asked to indicate the frequency with which they had ever engaged in a set of nine computer-related production activities from a choice of never, once or twice, three to six times, or more than six times. Activities queried included making a computer program, a website, digital art, a digital 2D or 3D model, a digital movie, an animation, a computer game, digital music, and a robot or other technology invention. Activities were presented as descriptions of possible products and/or potential software applications, such as “Made a publication or a newsletter (e.g. PageMaker, Word, Comic Life)” and “Designed a 2D or 3D model or drawing (e.g. CAD, ModelShop)” (For the full item, see Barron & Martin, 2014). The number of activities students reported any experience with were counted to create a breadth of digital production experience score.

### Technology in parent jobs

Students were asked to indicate the degree to which their parents’ jobs involved technology. First, they were asked if their parents used computers in their jobs from a choice of *Yes, No, I don’t know*; with mother/female guardian and father/male guardian queried separately (in the interest of space we refer to these two categories as *mothers* and *fathers* in the remainder of the paper). For those who responded *Yes*, a follow up question asked about the degree of occupational technology use from a choice of: *Occasionally as a productivity tool as part of a larger job (e.g. is an author and uses word processing, is a financial planner and uses record keeping software)* (occasional user) and *job is defined by the computer/technology; it is their primary tool (e.g. works as a Web designer, networking consultant, computer programmer, engineer, etc.)* (technology professional). Participants with unclear or incomplete answers were dropped from the analysis.

Silicon Valley is a technology hub with a growing concentration of the labor force in this industry. Eighty-eight percent of students for whom we have parent job information reported having at least one parent who was a technology professional. Over three quarters of fathers were classified as technology professionals, while mothers were more evenly distributed (Table 1). The predominance of fathers with technology-defined jobs (twice as many fathers as mothers for this sample) is reflective of a recent Silicon Valley workforce analysis, documenting persistent underrepresentation of women (John & Carnoy, 2017).

Table 1: Degree of technology in parent jobs for mothers and fathers

|                 | Degree of technology use in job |      |                 |      |                   |      |
|-----------------|---------------------------------|------|-----------------|------|-------------------|------|
|                 | No usage                        |      | Occasional user |      | Tech professional |      |
|                 | N                               | %    | N               | %    | N                 | %    |
| Mothers (n=316) | 75                              | 23.7 | 118             | 37.3 | 123               | 38.9 |
| Fathers (n=326) | 14                              | 04.3 | 58              | 17.8 | 254               | 77.9 |
| Both parents    | 6                               | 01.9 | 28              | 08.9 | 99                | 31.3 |

### Parent learning partner roles

The measure of parent learning partner roles is based on interview and ethnographic studies looking at roles played by adults that seem to be generative in young people’s digital production interests and hobbies (Barron *et al.*, 2009; Barron *et al.*, 2014). The item asked students if their parents (mother and father queried separately) ever played each of 10 roles to support their learning about computers and technology (Table 2). Roles include those that provide new opportunities for learning by connecting children with activities and resources and those that involve joint engagement as child and parent(s) work and learn together (Takeuchi & Stevens, 2011). In the Table 2, roles are organized within these two conceptual groups by frequency for this sample, from those most to least played.

Table 2: Survey descriptions of parent learning partner roles in youth technological learning

| Role                                 | Survey Item  |
|--------------------------------------|--|
| <i>Providing new opportunities</i>   |  |
| Financier of entertainment resources | Bought me entertainment related technology ( <i>like games, console</i> )  |
| Resource lender                      | Had things ( <i>like books, equipment, software</i> ) at the house that I use.   |
| Financier of educational resources   | Bought me things to support my computer activities and learning ( <i>like hardware, software, books, courses</i> )                                 |
| Learning broker                      | Looked for technology-related activities for me to do and/or signed me up for them ( <i>like classes, clubs, camps, etc.</i> )                     |
| Employer                             | Paid me to do something technical or on the computer for her/him.  |
| <i>Learning together</i>             |  |
| Teacher                              | Taught me how to do something on the computer ( <i>like typing, how to create a Web page, etc.</i> )   |
| Learner                              | I have taught them how to do something with the computer.  |
| Model                                | Let me watch how they do something ( <i>like turn on the computer, set up the printer</i> ) that I eventually learn how to do from observing them. |
| Non-tech consultant                  | Gave me advice on non-technical issues that helped me with my technology activities.   |
| Collaborator                         | Worked with me on a technology- or computer-related project ( <i>like built a robot together, worked on a Flash tutorial together</i> )            |

## Results

Consistent with our ethnographic study (Barron *et al.*, 2009) there was variability in terms of the number of learning partner roles played by mothers and fathers. Below we address three questions related to parent role variability in relation to child experiences and parent occupation.

### Did the number of roles played correlate with children's digital production experience?

We created a measure of the *diversity* of roles (number of roles played by mother or father, with a possible total of 10) and a measure of the *density* of roles (number of roles played by mother and father, with a possible total of 20). In this sample, students reported an average diversity of 6.90 roles (SD = 2.17) ranging from 1-10, and an average density of 11.12 roles (SD = 1.84), ranging from 1-20. There was a significant association between these measures and youth experiences with digital production. Pearson correlation indicated a positive association between digital production experience scores and diversity of parent roles ( $r(360) = .31, p < .001$ ) and between digital production experience scores and density of parent roles ( $r(360) = .32, p < .001$ ).

### Was diversity of roles played associated with level of parent technology use at work?

We also created measures of the *diversity* of learning partner roles each parent separately played for their children and looked at those measures by parent job classification (Figure 1). Almost twice as many roles were played by parents defined as technology professionals than by parents who did not use technology at work. An ANOVA with mother technology job focus as the between subjects factor revealed a significant effect of job on the number of roles played by mothers ( $F(2,313) = 31.51, p = <.001$ ). For mothers, these phenomena increased with the level of technology focus within their jobs, with tech professionals playing more learning partner roles than those who used technology occasionally in their jobs, and those mothers playing more roles than mothers who did work at all with computers. An ANOVA with father technology job focus as the between subjects factor also revealed a significant effect of job on the number of roles played ( $F(2,323) = 14.76, p = <.001$ ). Fathers who used technology at all at work played significantly more learning support roles than those who did not. For fathers, differences between technology professionals and occasional users and were smaller and not significantly different.

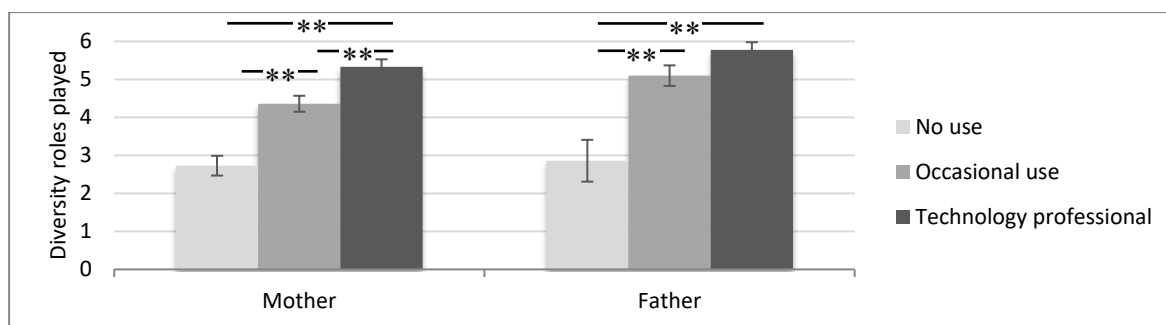


Figure 1. Number of learning partner roles played by parent (diversity) by parent job classification.

### Did the kinds of roles played by parents differ by parents with technology jobs?

Analysis reveals striking differences in roles played by parents with more and less technology focus in their professional life. For both mothers and fathers, using technology at work significantly increased the likelihood of engaging in every learning partnership role with the exception of parent as learner. Roles such as modeling, collaborating, and teaching were much more likely when parent jobs required the use of technology.

#### Mothers or female guardians

Mothers whose work was not at all defined by computers and technology (no usage) were less likely to play every role queried than those who did (occasional users and technology professionals) except for learning from their child and employing them to help with technology issues (Table 3). In terms of those roles that may provide new extended opportunities for learning, mothers who used technology at all in their jobs were more likely to finance equipment. This suggests differing levels of technology tools in the house and possibly financial differences between families. Mothers classified as technology professionals were more likely than both other groups to lend or share their own resources with their children and were more likely than mothers who did not use technology in their jobs to broker learning events for their children. In terms of learning together, roles that require some parent technology knowledge or expertise (teacher and model) were significantly different across all three groups, with likelihood increasing with increased technology job focus. Mothers classified as technology professionals were four times as likely as those who did not use technology in their work to teach their children about computers and technology and to model technical and computational practices, and three times as likely to collaborate with their children. Mothers who were technology professionals were more than twice as likely than the occasional users to collaborate with their children.

Table 3: Roles played by mother as a function of mother's use of technology at work

| Roles                              | Degree of technology use in job |                          |                            | $X^2$ | $p$   |
|------------------------------------|---------------------------------|--------------------------|----------------------------|-------|-------|
|                                    | No usage<br>n=75                | Occasional user<br>n=118 | Tech professional<br>n=123 |       |       |
| <i>Providing new opportunities</i> |                                 |                          |                            |       |       |
| Financier (entertainment)          | 44.0% <sup>a</sup>              | 63.6% <sup>b</sup>       | 72.4% <sup>b</sup>         | 16.08 | <.001 |
| Resource lender                    | 25.3% <sup>a</sup>              | 40.7% <sup>a</sup>       | 63.4% <sup>b</sup>         | 29.27 | <.001 |
| Financier (learning)               | 24.0% <sup>a</sup>              | 41.5% <sup>b</sup>       | 49.6% <sup>b</sup>         | 12.75 | .002  |
| Learning broker                    | 30.7% <sup>a</sup>              | 39.8% <sup>ab</sup>      | 55.3% <sup>b</sup>         | 12.61 | .002  |
| Employer                           | 06.7% <sup>ab</sup>             | 18.6% <sup>b</sup>       | 08.1% <sup>a</sup>         | 08.86 | .012  |
| <i>Learning together</i>           |                                 |                          |                            |       |       |
| Teacher                            | 18.7% <sup>a</sup>              | 52.5% <sup>b</sup>       | 76.4% <sup>c</sup>         | 62.65 | <.001 |
| Learner                            | 81.3% <sup>ab</sup>             | 86.4% <sup>b</sup>       | 72.4% <sup>a</sup>         | 07.57 | .023  |
| Model                              | 13.3% <sup>a</sup>              | 39.0% <sup>b</sup>       | 57.7% <sup>c</sup>         | 38.30 | <.001 |
| Non-tech consultant                | 17.3% <sup>a</sup>              | 38.1% <sup>b</sup>       | 41.5% <sup>b</sup>         | 13.11 | .001  |
| Collaborator                       | 12.0% <sup>a</sup>              | 15.3% <sup>a</sup>       | 35.8% <sup>b</sup>         | 20.74 | <.001 |
| Teaching and learning              | 16.0% <sup>a</sup>              | 43.2% <sup>b</sup>       | 52.8% <sup>b</sup>         | 26.82 | <.001 |

Items in a row that share a common superscript or that do not have superscripts do not differ significantly.

We were especially interested in parents playing the bidirectional role of both learner and teacher, considering

it a unique look into potentially generative interactions between parent and child as they share knowledge. In this sample, 55% of students reported at least one parent who both learned from and taught them, and 19% reported this to be true for both parents. There were no differences between occupation groups in terms of the frequency of fathers who played this bidirectional role but there were for mothers (Tables 3 and 4). Mothers who used technology in their jobs were more likely to have a teaching and learning relationship. It is of note that the majority of mothers in all groups are playing the learning role for their child, but this was not true for fathers.

### Fathers or male guardians

Father occupation groups did not differ in terms of brokering new learning activities for their children, but they did provide material resources differently (Table 4). Fathers who did not use technology at all in their jobs were less likely than the other groups to finance technology resources for their child than the other groups. Technology professionals not only purchased more learning resources for their children but were also more than twice as likely to lend their own resources than those fathers who did not use technology in their jobs. In terms of learning together, technology professionals were more likely than both other groups to teach their children, and this was true for occasional users compared to fathers that did not use technology at all in their job. On the other hand, the technology professionals were less likely than fathers who did not use technology at all in their jobs to learn from their child. Fathers who did not use technology in their jobs were more than twice as likely to learn from their children than fathers who were technology professionals. A similar pattern was observed for modeling and collaborating.

**Table 4: Roles played by father as a function of father's use of technology at work**

| Roles                              | Degree of technology use in job |                          |                            | $\chi^2$ | <i>p</i> |
|------------------------------------|---------------------------------|--------------------------|----------------------------|----------|----------|
|                                    | No usage<br>n=14                | Occasional usage<br>n=58 | Tech professional<br>n=254 |          |          |
| <i>Providing new opportunities</i> |                                 |                          |                            |          |          |
| Financier (entertainment)          | 42.9% <sup>a</sup>              | 77.6% <sup>b</sup>       | 82.7% <sup>b</sup>         | 13.45    | .001     |
| Resource lender                    | 28.6% <sup>a</sup>              | 63.8% <sup>ab</sup>      | 70.5% <sup>b</sup>         | 11.06    | .004     |
| Financier (learning)               | 28.6% <sup>a</sup>              | 53.4% <sup>ab</sup>      | 66.5% <sup>b</sup>         | 10.68    | .005     |
| Learning Broker                    | 00.0%                           | 24.1%                    | 26.8%                      | 05.09    | .078     |
| Employer                           | 14.3%                           | 12.1%                    | 13.4%                      | 00.09    | .958     |
| <i>Learning together</i>           |                                 |                          |                            |          |          |
| Teacher                            | 35.7% <sup>a</sup>              | 79.3% <sup>b</sup>       | 91.7% <sup>c</sup>         | 40.93    | <.001    |
| Learner                            | 78.6% <sup>a</sup>              | 46.6% <sup>ab</sup>      | 35.8% <sup>b</sup>         | 11.58    | .003     |
| Model                              | 28.6% <sup>a</sup>              | 67.2% <sup>b</sup>       | 79.1% <sup>b</sup>         | 20.19    | <.001    |
| Non-tech consultant                | 21.4% <sup>a</sup>              | 39.7% <sup>a</sup>       | 52.0% <sup>b</sup>         | 07.15    | .028     |
| Collaborator                       | 07.1% <sup>a</sup>              | 46.6% <sup>b</sup>       | 59.1% <sup>b</sup>         | 16.27    | <.001    |
| Teaching and Learning              | 21.4%                           | 32.8%                    | 33.5%                      | 00.87    | .647     |

Items in a row that share a common superscript or that do not have superscripts do not differ significantly.

## Discussion and future directions

Learning to use technology as a tool of creative production has been recognized as a key capacity that will help position young people for work, civic engagement, and future learning. Understanding how these capacities develop is critical for tracking variability in access to learning opportunities provided at home, in schools, or in virtual or online communities. In this paper we provided findings that stress the importance of interactions between people as *enactive* elements of the system and consider those interactions through the lens of *extended* environments, in this case parent jobs that influence family learning within a child's broader learning ecology. Describing and differentiating supportive roles played by parents and being able to quantify roles played in social environments has implications for measuring learning networks and corresponding design-based interventions. In this paper we focused on learning partnerships at home, asking whether parents who worked with computers and technology in their jobs played more roles for their children than parents who did not work in the field or use technology in their work at all. Our findings indicated that the diversity of roles experienced by young people was significantly related to parent use of technology at work.

Our Silicon Valley sample is from a highly unique community in terms of its technological expertise and capital, and specific patterns found here may not transfer to other locations or be broadly representative of the

U.S., although as shared in the introduction, there is variability even in this highly concentrated area. It would be productive to know more about the types of jobs that are represented in this sample and the expertise they require. Analyses of technical occupations suggest that jobs vary along multiple dimensions and we would expect that what parents know and might possibly share would vary accordingly. Beyond content knowledge, technical expertise is a form of cultural capital that includes values, beliefs identities, and access to knowledge networks. It may be the case that some regions or occupations provide access to specialized technically relevant social capital through networks and learning partnerships that can support knowledge development (Bourdieu,1986). Recent quantitative data on the geographical histories of patent holders suggests that regional social networks are critical for surfacing and sustaining talent (Bell, Chetty, Jaravel, Petkova, & Van Reenen, 2017). Andres, a learner described in the introduction to this paper, shared his knowledge in contributing to his church community, using Adobe design products to create a song guide that scrolled lyrics projected on a large screen. Jonathon’s consulting business was utilized by teachers at his school. From an equity perspective it is imperative to document the conditions that allow for and constrain the development and recognition of innovation of all types. Combining geographical analyses with more nuanced ethnographic research would set the stage for future experimental intervention-oriented research intended to help democratize the potential of all young people to innovate and share their innovations more broadly. As a first step, we intend to refine our retrospective survey questions to carry out a nationally representative survey that could see how these patterns play out or not across a larger and more diverse sample of youth and/or parents. Gender is also an area for more research, although in this sample there were no significant differences between boys and girls in this sample in terms of diversity or density of parent roles played or their own digital production experience. There were differences between mothers and fathers in the proportion engaged in technology focused jobs and in the frequency of roles. For example, mothers were more likely than fathers to broker learning opportunities for their children. In other analysis not reported here we show that the broker role is very important in this sample—students who report parents who play the learning broker role also report learning outside of school about a greater number of technology-related topics.

In closing, our findings underscore the value of deepening our understanding of informal family-based learning partnerships as a contributor to the technical expertise of young people. This is a particularly important goal for addressing issues of innovational equity (Barron, 2004). A well-documented lack of diversity in those who are lead designers is increasingly recognized as significant challenge and a threat to creativity and innovation in the technology industry. To the extent that we have a more homogenous group of professionals imagining and building future tools we fail to capitalize on diversity of perspectives, ultimately limiting potential solutions. It is generally agreed that this is a multidimensional problem that includes early gaps in experience, gender and racial stereotyping of technical work, and workplaces that create climates that suppress rather than invite contributions from all (e.g. Google & Gallup, Inc., 2016; Williams, Li, Rincon, & Finn, 2018). Our findings suggest that by supporting all parents as learning partners we might increase the likelihood of engaging a more diverse group of young people in activities that will position them to be future designers in service of their own goals, their communities’ needs, or their workplaces. Given that families are central sites for learning interactions, working to help expand their opportunities to co-learn may be one of the more important things we can do to fulfill the potential promise of technology as a resource for the greater good. At the same time, to support more equitable robust learning ecologies, it is essential that schools, libraries, community organizations and other public institutions provides access to opportunities to complement the important work that families do.

## References

- Barron, B. (2006). Interest and self-sustained learning as catalysts of development: A learning ecologies perspective. *Human Development*, 49, 193-224.
- Barron, B. (2004). Learning ecologies for technological fluency: Gender and experience differences. *Journal of Educational Computing Research*, 31(1), 1-36.
- Barron, B. (2010). Conceptualizing and tracing learning pathways over time and setting. In W. R. Penuel, & K. O'Connor (Eds.), *Learning research as a human science. National Society for the Study of Education Yearbook*, 109(1), 113-127.
- Barron, B., Gomez, K., Pinkard, N. and Martin, C.K. (2014). *The Digital Youth Network: Cultivating digital media citizenship in urban communities*. Boston, MA: MIT Press.
- Barron, B. & Martin, C.K. (2016). Making matters: A framework for assessing digital media citizenship. In K. Peppler, Y. Kafai, & E. Halverson (Eds.) *Makeology (Vol 2): Makers as learners* (pp. 45-72). New York: Routledge.
- Barron, B., Martin, C.K., Takeuchi, L., Fithian, R. (2009). Parents as learning partners in the development of technological fluency. *International Journal of Learning and Media*, 1(2), 55–77.

- Barron, B., Walter, S., Martin, C.K., and Schatz, C. (2010). Predictors of creative computing participation and profiles of experience in two Silicon Valley middle schools. *Computers and Education*, 54(1), 178-189.
- Bell, A. M., Chetty, R., Jaravel, X., Petkova, N., & Van Reenen, J. (2017). Who Becomes an Inventor in America? The Importance of Exposure to Innovation (No. w24062). National Bureau of Economic Research.
- Bourdieu, P. (2011). The forms of capital. In I. Szeman & T. Kaposy (Eds.), *Cultural theory: An anthology* (pp. 81-93). Malden, MA: Wiley-Blackwell.
- Ching, D., Santo, R., Hoadley, C., & Pepler, K. (2015). On-ramps, lane changes, detours and destinations: Building connected learning pathways in Hive NYC through brokering future learning opportunities. White paper for the HIVE Research Lab. Document retrieved from: <http://hivenyc.org>
- Google Inc. & Gallup Inc. (2016). Diversity Gaps in Computer Science: Exploring the Underrepresentation of Girls, Blacks and Hispanics. Retrieved from <http://goo.gl/Pg34aH>
- Gutiérrez, K. D., & Rogoff, B. (2003). Cultural ways of learning: Individual traits or repertoires of practice. *Educational Researcher*, 32(5), 19-25.
- Horrigan, John B. (2016). Digital Readiness Gaps. Pew Research Center. Retrieved from <http://www.pewinternet.org/2016/09/20/2016/Digital-Readiness-Gaps/>
- Ito, M., Horst, H., Antin, J., Finn, M., Law, A., Manion, A. et al. (2009). *Hanging out, messing around, and geeking out: Kids living and learning with new media*. Cambridge, MA: MIT Press.
- John, J. & Carnoy, M. (2017). Race and gender trends in computer science in the Silicon Valley from 1980-2015. Retrieved from [https://cepa.stanford.edu/sites/default/files/JohnCarnoy\\_Sept2017\\_0.pdf](https://cepa.stanford.edu/sites/default/files/JohnCarnoy_Sept2017_0.pdf)
- Livingstone, S., Ólafsson, K., Helsper, E., Lupiáñez-Villanueva, F., Veltri, G., & Folkvord, F. (2017). Maximizing opportunities and minimizing risks for children online: The role of digital skills in emerging strategies of parental mediation. *Journal of Communication*, 67, 82-105.
- Louw, M., Barbuto, N., & Crowley, K. (2016). Designing Learning Pathways in a complex learning ecology: A research-practice partnership focused on parent brokering. In B. DiSalvo, J. Yip, E. Bonsignore, & C. DiSalvo (Eds), *Participatory Design for Learning: Perspectives from Research and Practice* (pp. 93-112). New York, NY: Routledge.
- Martin, C.K., Barron, B., Stringer, D. & Matthews, J. (2014). Patterns of engagement: How depth of experience matters. In B. Barron, K. Gomez, N. Pinkard, and C.K. Martin *The Digital Youth Network: Cultivating Digital Media Citizenship in Urban Communities*. Boston, MA: MIT Press.
- Peng, G. (2017). Do computer skills affect worker employment? An empirical study from CPS surveys. *Computers in Human Behavior*, 74, 26-34.
- Rogoff, B. (1998). Cognition as a collaborative process. In W. Damon (Series Ed.) & D. Kuhn & R.S. Siegler (Volume Eds.), *Handbook of Child Psychology, Vol. 2: Cognition, perception and language* (pp. 679-744). New York: Wiley.
- Rogoff, B. (2003). *The cultural nature of human development*. Oxford, UK: Oxford University Press.
- Takeuchi, L., & Stevens, R. (2011). The new coviewing: Designing for learning through joint media engagement. New York, NY: The Joan Ganz Cooney Center at Sesame Workshop.
- Whiting, B., & Whiting, J. (1970). Methods for observing and recording behavior. In R. Naroll & R. Cohen (Eds) *A handbook of method in cultural anthropology, Vol 2* (pp. 282-315). Garden City, N.Y: Natural History Press.
- Williams, J.C., Li, S., Rincon, R., & Finn, P. (2018). Climate control: Gender and racial bias in engineering? Center for Worklife Law & Society of Women Engineers. Retrieved from <https://research.swe.org/climate-control/>
- Yuen, A. H., Park, J., Chen, L., & Cheng, M. (2018). The significance of cultural capital and parental mediation for digital inequity. *New Media & Society*, 20(2), 599-617.

## Acknowledgements

This research was funded through the LIFE (Learning in Informal and Formal Environments) Science of learning Center, an NSF-funded (REC-354453) effort seeking to understand and advance human learning through a simultaneous focus on implicit, informal, and formal learning, thus cultivating generalizable interdisciplinary theories that can guide the design of effective new technologies and learning environments. Writing and analyses were supported by the TELOS (Technology for Equity in Learning Opportunities) initiative at Stanford University.

# From Material Objects to Social Objects: Researching the Material-Dialogic Spaces of Joint Attention in a School-Based Makerspace

Kristiina Kumpulainen and Anu Kajamaa  
kristiina.kumpulainen@helsinki.fi, anu.kajamaa@helsinki.fi  
University of Helsinki

**Abstract:** In this paper, we investigate the material-discursive spaces of joint attention between students and their teachers in a school-based makerspace. Drawing on sociocultural theorizing, Bakhtinian inspired notion of the “dialogic space”, and material-discursive onto-epistemology, we introduce the concept of *social object* to explain how material objects can turn into joint attention and meaning-making in ongoing interaction. Video recordings (75 hours) of 94 students’ aged 9 to 12 years old and their teachers’ interactions in the makerspace were analyzed. The study shows how the material objects of the makerspace were enacted into social objects via joint attention *about* the objects, *around* the objects and *with* the objects. The study demonstrates the need to rethink the meaning and role of material objects in novel educational makerspaces for establishing and negotiating joint attention, mediating learning opportunities and tensions.

**Keywords:** social object, material-dialogic space, joint attention, educational, makerspace, engagement and learning

## Introduction

Emerging digital technologies and infrastructures are changing educational practices with new possibilities and tensions for interaction and learning (Kumpulainen, Mikkola, & Rajala, 2018; Rasmussen & Ludvigsen, 2010; Säljö, 2010). Hence, continual investigation of how people interact with new cultural tools, the meanings they give to these tools, and how these new tools are hybridized with and across people, existing tools and practices is essential. In this paper, we focus on how tools, in specific, material objects of a novel educational makerspace are enacted in the material- discursive spaces of joint attention between students and their teachers. We argue, that at present, little attention has been directed to the meaning and role of material objects in mediating students’ joint attention, engagement and learning in novel, material-rich educational makerspaces, encompassing the generative use of both traditional and more recent technologies, such as, 3D-printers, electronics, and design apps. Our study holds that that joint attention, that is, the human capacity to coordinate actions and attention with others on an object (Tomasello, 2000) is crucial for collaborative meaning-making and problem-solving, and it creates the foundation for social, cognitive and emotional learning (Bruner, 1995). It is hence important to generate new knowledge on the ways in which material objects of educational makerspaces can turn into joint attention in ongoing interaction with possibilities and tensions for students’ engagement and learning.

Educational makerspaces prescribe a constructionist model of student-centered pedagogies in which students can work on personally meaningful Science, Technology, Engineering, Arts and Mathematics (STEAM) design projects, where they can make choices about their activities: simultaneously navigating through several fields of knowledge and using novel technologies that enable them to externalize, share and build ideas into concrete material objects (Halverson & Sheridan, 2014; Honey & Kanter, 2013; Pepler, Halverson, & Kafai, 2016). Emerging research around makerspaces suggests that they can foster students’ creative problem-solving, STEM learning and 21st century skills, including digital literacy (see e.g. Honey & Kanter, 2013; Kafai, Fields, & Searle, 2014; Lindtner, 2014; Martinez & Stager, 2013). However, more research is deemed necessary to generate a more comprehensive understanding about the educational potential of makerspaces.

We know from earlier body of research that achieving and maintaining joint attention, collaboration and knowledge co-construction are important but demanding activities on their own right (Bruner, 1995; Ludvigsen, 2012). This complexity related to achieving and negotiating joint attention and meaning is likely to increase in educational makerspaces in which students work on complex interdisciplinary design challenges that encompass not only processes of creating specific material objects supported by a wide range of technologies and media, but also cognitive, emotional, relational and cultural processes surrounding the use and construction of material objects (Kumpulainen, Kajamaa, & Rajala, 2018). There are hence important insights to be gained from researching the material-dialogic spaces of joint attention between students and their teachers in educational makerspaces as to identify learning opportunities and tensions that can inform further educational design and support.

## Conceptual background

Drawing on sociocultural theorising (Vygotsky, 1986, 1997), Bakhtinian inspired notion of the “dialogic space” (Wegerif, 2011), and material-discursive onto-epistemology (Barad, 2003, 2007), in this study, we introduce the concept of *social object* to explain how material objects can turn into joint attention and meaning-making in ongoing interaction in an educational makerspace. In our conceptualising, we do not automatically view material objects as social, but only then when they are integrated and taken up in the material-discursive activity. We consider social objects as transactional, facilitating joint attention and productive exchanges among those who encounter them (Knorr-Cetina, 1997; Simon, 2010).

Vygotsky and his co-investigators seminal sociocultural work on tool-mediation, widens our understanding of how humans, through tool-mediated activities, transform the environment in which they live (Vygotsky 1986; Vygotsky & Luria 1994). Vygotsky saw material-semiotic tools as constitutive of human activity and a prominent driving force for the development of human mind and culture (Vygotsky, 1986, 1997). For Vygotsky, language was the tool of tools, however, he did not undermine the mediational role of material objects (i.e. tangible tools and artefacts) for human learning and development. In fact, the centrality of materiality in human activity advocated by the sociocultural theories remind us of how social action and tangible and conceptual (signs and language) tools are intertwined (see also Ingold, 2010; Mäkitalo, 2011). Material objects are particularly important as they carry cultural knowledge both for the individual and collectives about their history, purpose and use, including values and ideologies. Furthermore, material objects can act as mediational means for personal and/or collective remembering (Cole, 1996; Wertsch, 2002), and enhance students’ joint attention, engagement, learning and development (Wartofsky, 1979; Paavola et al., 2004; Kumpulainen, Kajamaa, & Rajala, 2018).

Sociocultural studies often refer to the Bakhtinian (1986) notion of a ‘dialogic space’ as a specific communicative event that evidences exploration, problematization and elaboration of diverse views and understanding in reasoned dialogue (Wegerif, 2011; Mercer et al., 2010). We extend this idea by introducing the notion of material-dialogic space. Following Hetherington & Wegerif (2018), we hold that a material-discursive emphasis can be used to extend the original dialogic account of the production of meaning in a way that it re-focuses attention on the multiple ‘voices’ of the material in co-mediating students’ joint attention, engagement and learning. In the material-dialogic spaces, the material and discursive are seen as entangled and intra-acting, with phenomena performed as a result of ongoing intra-actions. Hence, the notion of intra-action underscores the mutually constitutive, entangled nature of matter and meaning (Barad, 2007). These intra-actions and performances generating from them, to which both people and materials contribute, co-create material-dialogic spaces with implications to ontological, epistemic, social, and ideological processes.

In this paper, we empirically investigate the material-dialogic spaces that constitute joint attention, engagement and learning between students and their teachers in an educational makerspace. To this end, we ask; How do material objects turn into and function as social objects in the material-dialogic spaces of joint attention? and How do social objects create opportunities and tensions for students’ engagement and learning?

## Empirical study

The empirical data stem from a city-run comprehensive school. The school follows the national core curriculum, which has been defined locally. The local curriculum of the school stresses design learning, which is considered to enhance students’ creative problem-solving skills across the curriculum. The school strives for learner-centeredness and for innovations in learning and teaching as expressed in its local curriculum. As a response, the school has recently introduced a new educational makerspace, the FUSE Studio as part of its elective courses.

The FUSE Studio is an educational makerspace, ‘a choice-based digital infrastructure for STEAM learning’ (see Stevens & Jona, 2017). The technological infrastructure of the FUSE Studio offers students with different STEAM challenges that ‘level up’ in difficulty like video games. The challenges include Spaghetti Structures, Jewellery Designer, Robot Obstacle Course, Keychain Customiser, Electric Apparel, Coaster Boss and Solar Roller. The challenges are accompanied by various tools, such as computers, 3D printers and other materials (e.g., a foam rubber, a marble, tape and scissors), as well as instructions on how to process the challenges.

Each FUSE challenge is designed to engage students in different STEAM topics and skill sets. The challenges have been carefully structured to introduce students to new ideas and to support them through more complex iterations of those ideas. Students can choose, based on their own interests, which challenges they want to work on, when and with whom. They can choose to work alone, in pairs or in small groups. There is no formal grading or assessment by teachers. Instead, by using photos, video or other digital artefacts, students can document their completion of a challenge, and the completion unlocks the next challenge.

Altogether, 94 students aged 9-12 years old and their teachers (six male and two female teachers) took part in the study. Due to the elective nature of the course, the groups consisted of students from several classes.



Group 1 consisted of 32 students (22 boys and 10 girls), Group 2 consisted of 30 students (19 boys and 11 girls) and Group 3 consisted of 32 students (19 boys and 13 girls). Each group was supported by two to four teachers and teaching assistants. At the beginning of the autumn, each group had one 45-minute FUSE session a week. Later in the autumn, each session was extended to 60 minutes.

## Methods

The primary data comprise 75 hours of video recordings of students carrying out design challenges in the FUSE Studio makerspace. The recordings were collected intermittently every week over a period of one semester by our team of researchers with four cameras in total. The episodes significant to our research focus in this paper, were transcribed and analyzed.

The video data were analysed using interaction analysis methods that took account of verbal, visual and material conduct in the students' and teachers' activity in the makerspace (Jordan & Henderson, 1995). The data were approached inductively by first approaching the video corpus as a whole and then focusing on selected events in greater depth (Derry et al., 2010). In particular, we were interested in those moments in the video data that gave evidence of the emergence of material-dialogic spaces of joint attention among the students and their teachers. Central to our analysis is a multi-step, multi-phase recursive analysis process, following the ethnographic logic of inquiry (Canstaneira, Green, & Dixon, 2009). For this, we watched the videos several times in an attempt to identify the regularities and patterns in the data source (Roth, 2005). Selecting interactional episodes evidencing the emergence and negotiation of joint attention (see also Gresalfi, Martin, Hand, & Greeno, 2009), we were able to trace how material objects turned into and began to function as social objects, mediating students' engagement and learning opportunities.

## Major findings

Our findings reveal three distinct, yet often overlapping, interactional processes during which the material objects of the makerspace developed into social objects in the material-dialogic spaces of joint attention between the students and their teachers. Namely, the study shows how the material objects of the makerspace were turned into social objects via joint attention and social interaction *about* the objects, *around* the objects and *with* the objects. We shall illuminate these three dominant manifestations of social objects in the vignettes below. These vignettes also make visible how the material dialogic spaces of joint attention created opportunities and tensions for students' engagement and learning in the makerspace.

### Vignette 1: Hey, what's this 'hole thing'?

This vignette illuminates a material-dialogic space of joint attention in which the students are wondering and experimenting about the use and functioning of a FUSE Studio design software, supported by the teacher. The vignette demonstrates the students' sense-making about the material objects in the educational makerspace. Here, it becomes clear how engagement with and learning to use the advanced technological tools of the FUSE Studio makerspace is pivotal for gaining access and authority in making activities in this space.

1. Student 1/Mel: Hey, what is this hole thing?
2. Teacher: What thing?
3. Student 1/Mel: Hole. That hole.
4. Teacher: Yeah you're supposed to make a hole there.
5. Student 1/Mel: What hole?
6. Teacher: Or what?
7. Student 1/Mel: No, I mean what is that hole?
8. Student 2/Anne: Yeah, what does it do?
9. Teacher: Is it that it turns transparent for a while so you can... Click on it again so it goes into the hole-mode. [girl clicks]
10. Teacher: Yeah so it shows that you can, it turns transparent so you can see through it. If you have that kind of situation that your planning requires you to be able to look through it.
11. Student 1/Mel: What do we do now that we're ready?
12. Student 2/Anne: We should probably look at the video.
13. Teacher: Look at the directions, I can't remember by heart.

14. Student 1/Mel: You're guiding it now Anne, I can't be using that thing [the laptop] the whole time.
15. Student 2/Anne: Okay. So, let's continue. [clicks open the directions video on the website]
16. Student 1/Mel: This is so slow... We have done that already.
17. Student 2/Anne: Okay, new video. [girls continue watching, teacher leaves]

In this example, two girls are sharing a laptop and working together on a design challenge the Keychain Customizer. They are designing a model of a keychain but are unsure of the software Tinkercad's commands. For that reason, they ask for the teacher's help. The teacher suggests that they try out the hole-command (line 9). He does not demonstrate using the command himself but encourages the students to do it. Then, the teacher explains what the command does and why one might use it (line 10). When the students ask what to do next, the teacher guides the students to look at the directions, like Anne suggests herself (12). By asking the students to do so, the teacher confirms a practice advocated by the educational makerspace that students should try to use other resources before asking for a teacher's help. Consequently, the material-dialogical space of the students' joint activity is expanded to involve the set of resources found in the website to support the students' independent engagement with the challenges and the associated technology. Therefore, the teacher points to the rules of the task and refocuses the students on it. This makes Mel assign the computer turn to Anne (line 14) who agrees (line 15). When Mel and Anne start looking for the directions, the teacher stays and listens to their discussion. When they are refocused on the task, the teacher leaves.

## Vignette 2: The stickiness of social objects



Figure 1. 3D printer as a social object.

Our analysis of the data reveals that often times the social objects of the educational makerspace functioned as 'a glue' that brought the participants, both students and teachers, together *around* the material objects of the makerspace to jointly observe, wonder, discuss and/or share, as demonstrated by the above picture (see Figure 1). Often times, in these situations joint attention was achieved by gesture and other non-verbal modes of interaction. These material-discursive spaces were filled with silence and everyone intensively observing what was happening whilst the technology (such as 3D printer) or a human (a student or teacher) constructed or developed something with the material objects. We could also identify conversations about the matter and its meaning situated in the present, the students explaining what was happening or giving instructions about alternative ways of working or using other material objects than those available in the makerspace. Excitement and interests, and emotional

engagement in general were made visible by non-verbal and verbal communication in the material-dialogic spaces around social objects.

### Vignette 3: Messing around with ones' own social objects

Our third vignette illuminates an alternative or, in fact, a competing material-dialogic space of joint attention constructed into being in the FUSE Studio makerspace *around* a material object. In this case, the student's own mobile phone. Here, the students are engaged in another activity they have found more meaningful instead of working on a FUSE design challenge. That is, they are playing around the Clash Royale game with their mobile phones.

[teacher walks into hallway to check on boys]

1. Student 1/Pekka: They're just playing...
2. Teacher Greg: Hey, what game do you have going on here?
3. Student 1/Pekka: They're playing Clash Royale...
4. Teacher Greg: Hey, put Clash Royale in your pocket and games away.

In this episode, the teacher walks into the hallway where a group of boys are working on a Coaster Boss design challenge. One of the students Pekka in the group responds immediately to the teacher's presence by explaining why their work is not coming along ("They are just playing", line 1). The teacher has a strong stance with his arms on his waist which can also be interpreted as a nonverbal sign of authority as he asks the students to stop playing the game. (line 4). The students obey the teacher's request and continue working on the FUSE design challenge. Overall, this example demonstrates the co-presence of at least two material-dialogic spaces that are performed in parallel, that is, working on the design challenge and playing the digital game. The co-presence of multiple material-dialogic spaces is enhanced by the fact that the online world and students' mobile phones are also commonly used in the FUSE Studio makerspace for the design challenge activities and their documentation. The teacher's actions in the vignette can be seen as an attempt to contain and constrain the dialogical-material space of the students' activity. However, it also demonstrates how digital tools define and alter the nature of the material dialogic space and the objects of joint attention.

### Vignette 4: Making a Dream Home

Vignette 4 demonstrates how material objects intra-acted in the material-discursive spaces of joint attention as explicit mediational means to explain *with*. In this example, the students are working on a Dream Home design challenge.

1. Student 1/Tara: I would like to turn this, so I could get to the other side.
2. Student 2/Hanna: Me too, because I don't even know how to get there.
3. Teacher John: Well wait, let's see who is furthest along in Dream Home. Eric and Ian, have you rotated the angles there, so you can get to the other side of the house?
4. Student 3/Rick: I have!
5. Teacher Greg: Hold down the mouse's button and then spin.
6. Teacher John: Okay, Rick can come instruct.
7. Student 3/Rick: [comes over to the girls] What?
8. Student 2/Hanna: How on earth do you turn this?
9. Teacher John: Hold down the mouse and...
10. Student 3/Rick: What did you want to do?
11. Student 1/Tara: Rotate the angle.
12. Student 3/Rick: Take that and then...[Tara rotates]
13. Teacher John: Which one was it Rick? Why don't you show me too.
14. Student 3/Rick: This tool.
15. Teacher John: Oh!

Here, Tara and Hanna have asked the teacher for help in rotating the view, so they could see the whole house. The teacher's first response is to find other students to help (line 3). By asking other students to help, the teacher is encouraging relative expertise in which the students can act as experts on the challenges. Student Rick is eager to help and comes over to advise the girls. After this, the teacher asks Rick to show him how to do it as well (line 13). By doing so, the teacher indicates that it is acceptable that teachers do not always know what to do in all of the challenges. He also reinforces Rick as an expert of the challenge. Interestingly, other teacher Greg exclaims the instructions in the middle of the conversation (line 5), even though he is helping other students at the time. Teacher Greg is probably aiming to speed the helping process, but this is in conflict with the other teacher's intervention strategy and that of the pedagogical model of the FUSE Studio makerspace that advocates for relative expertise. Teacher John does repeat these instructions partly (line 9) but then lets the student Rick to help and explain it to the two girls.

## Conclusions and implications

Our findings that stem from our ongoing research work on an educational makerspace in a school context, make visible the nuanced ways in which the material objects of the educational makerspace were turned into and functioned as social objects in the interactions between and across the students and their teachers. Our findings reveal three distinct and overlapping interactional processes during which the material objects of the makerspace turned into social objects via joint attention and social interaction *about* the objects, *around* the objects and *with* the objects.

First, there were material-dialogic spaces in which joint attention was established via social interaction *about* the material objects themselves. This mode of joint attention in relating to material objects became evident especially when the habitual ways of engagement were disrupted, for example, by technological failures or discrepancy between means and ends of the activity (Kumpulainen, Rajala, & Kajamaa, in press). The problems in the technological infrastructure also created uncertainty among the teachers and challenged their role as authority as they did not always have control over the material objects themselves either. Secondly, we could also identify material- dialogic spaces in which joint attention was enacted and negotiated *around* the material objects. Our analysis suggests that the contemporary pedagogical and digital infrastructure engender dynamically shifting and expansive material-dialogic spaces that create new possibilities and tensions for students' engagement and learning opportunities. For example, the group configurations in the FUSE studio makerspace are flexible and the students are invited to work across groups to help each other. Similarly, digital tools offer vast possibilities for expanding the scope of the activity and dialogue (Kumpulainen, Mikkola, & Rajala, 2018). This imposes tensions for joint attention among students and between students and their teachers. Thirdly, we depicted material-dialogic spaces in which joint attention was established and maintained *with* the material objects. Here, the meaning potential of material objects in co-ordinating joint action and attention was pivotal. In all, it was in these sociomaterial and dialogic encounters that constituted the material-dialogic spaces of joint attention, when the material objects of the educational makerspace turned into social objects and mediated the students' engagement and learning opportunities.

The study reveals a dynamic interplay of multiple voices emerging in the material-dialogic spaces of joint attention. The inter-animation of voices that were performed into being in ongoing interactions evidence delicate and at times strong power relationships in the positioning of different voices, with consequences for engagement and learning opportunities. At times, the voices of material objects dominated the material-dialogic space of joint attention, and undermined the voices of others. At other times, it was the teacher's voice that become more authoritative from other voices, with opportunities and tensions for student-centered learning. We could also depict material-dialogic spaces that demonstrate joint reasoning and meaning making between the students and teachers, giving rise to relative expertise and enhancing the students' interest-driven creative activities and learning opportunities.

The results make visible how the material objects of the makerspace when turned into social objects of joint attention are important mediators of power and educational equity, making materiality a pivotal research focus for future studies in education. In our research not all the students were found to engage in interest-driven STEAM learning activities in the educational makerspace despite free choice of the design challenges they could work on. Our observations resonate with existing research that has pointed out how makerspaces hosted by various educational and cultural institutions often fail to attract and engage the broader population of young people in learning due to culturally biased materialities and activities (Barton, Tan & Greenberg, 2016; Peppler et al., 2016). Our research joins in with these concerns and calls for the quality and inclusivity of makerspaces and their materialities, and urges further investigation into novel, material rich educational spaces as they are related to creating democratic, equitable, and deep learning experiences for diverse student. Moreover, further research on

materiality is a pivotal research focus for future studies in education for promoting productive interaction in novel educational settings. In sum, the study demonstrates the need to rethink the meaning and role of material objects in novel educational makerspaces for establishing and negotiating joint attention, mediating learning opportunities and tensions.

## References

- Alexander, R. J. (2008). *Towards Dialogic Teaching: rethinking classroom talk* (4th edition). New York: Dialogos.
- Bakhtin, M. (1986). *Speech genres and other late essays*. Trans. Vern W. McGee. Austin, Texas: University of Texas Press.
- Barad, K. (2003). Posthumanist performativity: Toward an understanding of how matter comes to matter. *Signs*, 28(3), 801–31.
- Barad, K. (2007). *Meeting the Universe Half-Way: Quantum Physics and the Entanglement of Matter and Meaning*. Durham: Duke University Press.
- Barton, A. C., Tan, E., & Greenberg, D. (2016). The makerspace movement: Sites of possibilities for equitable opportunities to engage underrepresented youth in STEM. *Teachers College Record*, 119(6), 11-44.
- Bennett, J. (2010). *Vibrant Matter: A Political Ecology of Things*. Durham, NC: Duke University Press.
- Braidotti, R. (2013). *The posthuman*. Cambridge, UK: Polity Press.
- Canstaneira, M.L., Green, J., & Dixon, C. (2009). Interactional ethnography: an approach to studying the social construction of literate practices. *Linguistics and Education*, 11, 353-400.
- Bruner, J. (1995). From Joint Attention to the Meeting of Minds: An Introduction. In C. Moore & P. J. Dunham (Eds.), *Joint Attention. Its Origins and Role in Development*. New York: Psychology Press.
- Cole, M. (1996). *Cultural psychology: A once and future discipline*. Cambridge, MA: Harvard University Press.
- Deleuze, G. & Guattari, F. (1994). *What is philosophy?* New York, NY: Columbia University Press.
- Goodwin, C. (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics*, 32, 1489-1522.
- Gresalfi, M., Martin, T., Hand, V., & Greeno, J. (2009). Constructing competence: an analysis of student participation in the activity systems of mathematics classrooms. *Educational Studies in Mathematics*, 70, 49-70
- Dewey, J. (1933). *How We Think A Restatement of the Relation of Reflective Thinking to the Educative Process*. Boston, MA D.C. Heath & Co Publishers.
- Derry, S. J., Pea, R. D., Barron, B., Engle, R. A., Erickson, F., Goldman, R. & Sherin, B. L. (2010). Conducting video research in the learning sciences: Guidance on selection, analysis, technology, and ethics. *The Journal of the Learning Sciences*, 19(1), 3 – 53.
- Fenwick, T., & Landri, P. (2012). Materialities, textures and pedagogies: socio-material assemblages in education, *Pedagogy, Culture & Society* 20(1), 1-7, DOI: 10.1080/14681366.2012.649421.
- Halverson, E. R. & Sheridan, K. (2014). The maker movement in education. *Harvard Educational Review*, 84(4), 495-504.
- Hetherington, L. & Wegerif, R. (2018). Developing a material-dialogic approach to pedagogy to guide science teacher education. *Journal of Education for Teaching*, 44(1), 27-43.
- Ingold, T. (2010). The textility of making. *Cambridge Journal of Economics*, 34, 91–102.
- Jordan, B. & Henderson, A. (1995). Interaction analysis: Foundations and practice. *The Journal of the Learning Sciences*, 4(1), 39-103.
- Kafai, Y., Fields, D., & Searle, K. (2014). Electronic textiles as disruptive designs: Supporting and challenging maker activities in schools. *Harvard Educational Review*, 84(4), 532-556.
- Knorr-Cetina, K. (1997). Sociality with objects. *Social Relations in Postsocial Knowledge Societies. Theory, Culture and Society*, 14(4), 1-30.
- Kuby, C. R. & Roswell, J. (2017). Early literacy and the posthuman: Pedagogies and methodologies. *Journal of Early Childhood Literacy*, 17(3), 286-296.
- Ludvigsen, S. R. (2012). What counts as knowledge: Learning to use categories in computer environments. *Learning, Media and Technology*, 37(1), 40-52.
- Kumpulainen, K., Kajamaa, A., & Rajala, A., (2018). Understanding educational change: Agency-structure dynamics in a novel design and making environment. *Digital Education Review*, 33, 26–38.
- Kumpulainen, K., Rajala, A., & Kajamaa, A. (in press). Researching the materiality of communication in an educational makerspace: The meaning of social objects. In N. Mercer, R. Wegerif, & L. Major (Eds.), *The Routledge International Handbook of Research on Dialogic Education*. London: Routledge.

- Kumpulainen, K., Mikkola, A., & Rajala, A. (2018). Dissolving the digital divide: Creating coherence in young people's social ecologies of learning and identity building. In J. Voogt, G. Knezek, R. Christensen, & K.-W. Lai (Eds.), *Second international handbook of information technology in primary and secondary education*. New York, NY: Springer.
- Mercer, N., Hennessy, S., & Warwick, P. (2010). Using interactive whiteboards to orchestrate classroom dialogue. *Technology, Pedagogy and Education*, 19(2), 195-209.
- Mäkitalo, Å. (2011). Professional learning and the materiality of social practice. *Journal of Education and Work* 25(1), 59-78.
- Paavola, S., Lipponen, L., and Hakkarainen, K. (2004). Models of innovative knowledge communities and three metaphors of learning. *Review of Educational Research*, 74(4), 557–576.
- Peppler, K., Halverson, E., & Kafai, Y. B. (Eds.). (2016). *Makeology: Makerspaces as learning environments* (Vol. 1). New York: Routledge.
- Rasmussen, I., & Ludvigsen, S. (2010). Learning with computer tools and environments: A sociocultural perspective. *International handbook of psychology in education*, 399-435.
- Resnick, L. B., Asterhan, C., & Clarke, S. (Eds.). (2015). *Socializing intelligence through academic talk and dialogue*. Washington D.C.: American Educational Research Association.
- Roth, W.-M. (2005). *Doing qualitative research: Praxis of method*. Rotterdam
- Simon, N. (2010). The participatory museum. Accessed from: <http://www.participatorymuseum.org/read/>
- Stevens R. & Jona, K. (2017). Program Design. FUSE Studio - website. Retrieved May 20, 2017 from <https://www.fusestudio.net/program-design>
- Säljö, R. (2010). Digital tools and challenges to institutional traditions of learning: technologies, social memory and the performative nature of learning. *Journal of Computer Assisted Learning*, 26(1), 53-64.
- Taylor, C.A. (2016). Edu-crafting a cacophonous ecology: Posthumanist research practices for education. In C.A. Taylor & C. Hughes (Eds.), *Posthuman Research Practices in Education* (pp. 5-24). New York: Palgrave MacMillan.
- Tomasello, M. (2000). *The cultural origins of human cognition*. London: Harvard University Press.
- Vygotsky, L. S. (1986). *Thought and language*. Cambridge, MA: MIT Press.
- Vygotsky, L. S. (1997). The instrumental method in psychology. In R. Reiber & J. Wollock (Eds.), *The collected works of L. S. Vygotsky* (Vol. 3). *Problems of the theory and history of psychology* (pp. 85–89). London: Plenum Press.
- Vygotsky, L. S., & Luria, A. (1994). Tool and Symbol in Child Development. In R. Van Der Veer, & J. Valsiner (Eds.), *The Vygotsky Reader*. Oxford: Blackwell Publishers.
- Wartofsky, M. (1979). *Models, representation, and the scientific understanding*. Boston: Reidel.
- Wegerif, R. (2011). "Towards a Dialogic Theory of How Children Learn to Think." *Thinking Skills & Creativity* 6 (3): 179–190.
- Wertsch J.V. (1991). *Voices of the Mind: A Sociocultural Approach to Mediated Action*. Cambridge, MA: Harvard University Press.
- Wertsch, J. V. (2002). *Voices of collective remembering*. Cambridge University Press.
- Wells, G. (1999). *Dialogic inquiry. Towards a sociocultural practice and theory of education*. USA: Cambridge University Press.

## Acknowledgments

The research work reported in this chapter has been financially supported by Learning by Making: The educational potential of school-based makerspaces for young learners' digital competencies (iMake) project funded by the Academy of Finland (project no: 310790).

# Analyzing Students' Synergistic Learning Processes in Physics and CT by Collaborative Discourse Analysis

Caitlin Snyder, Vanderbilt University, caitlin.r.snyder@vanderbilt.edu  
Nicole Hutchins, Vanderbilt University, nicole.m.hutchins@vanderbilt.edu  
Gautam Biswas, Vanderbilt University, gautam.biswas@vanderbilt.edu  
Mona Emara, Damanhour University, mona.emara@damanhour.edu.eg  
Shuchi Grover, Looking Glass Ventures, shuchig@cs.stanford.edu  
Luke Conlin, Salem State University, luke.conlin@gmail.com

**Abstract:** The introduction of computational modeling into science curricula has been shown to benefit students' learning, however the synergistic learning processes that contribute to these benefits are not fully understood. We study students' synergistic learning of physics and computational thinking (CT) through their actions and collaborative discourse as they develop computational models in a visual block-structured environment. We adopt a case study approach to analyze students synergistic learning processes related to stopping conditions, initialization, and debugging episodes. Our findings show a pattern of evolving sophistication in synergistic reasoning for model-building activities.

## Introduction

Computation is at the forefront of 21st century education (Wing, 2006). Technological advancements are resulting in the introduction of a variety of computational tools and practices into Science, Technology, Engineering, and Mathematics (STEM) curricula through computational modeling, simulation, data analysis and visualization. Working in technology-enhanced environments also presents opportunities for collaborative learning and problem-solving. We aim to integrate computational thinking (CT) concepts and practices into STEM curricula (Sengupta, et al., 2013) so that learning of STEM and CT can be mutually supportive and help to develop important STEM practices like model-based reasoning, explanation, and argumentation. Computational modeling serves as an effective vehicle for science learning in K-12 (e.g. Basu, et al., 2016; diSessa, 2001; Wilensky & Reisman, 2006). Designed learning environments can support this mutually supportive integration (Basu, et al., 2013; Hutchins, et al., 2018), and such environments have helped students achieve significant learning gains in both STEM and CT concepts and practices (Basu, et al., 2017; Sengupta, et al., 2013). But the mechanisms and processes students employ to develop and apply synergistic learning skills are not well understood.

In this paper, we take on the challenge of understanding and unpacking students' synergistic learning processes while they develop computational models of scientific phenomena using a block-based programming language. Typically, computational modeling involves an iterative process of conceptualization, algorithmic design, implementation as a program, and testing and refining the program to generate a correct model. Students activities related to these processes can be analyzed using log data collected from computer-based learning environments. However, students' underlying reasoning mechanisms when invoking and applying these processes, the difficulties they face, and how they overcome them are hard to unpack from their logged activities. To better understand these mechanisms, we ran a pilot study with students who worked in groups of two or three on model building and problem-solving tasks in physics. We used a screen recording system that simultaneously captured students' interactive dialog as they worked on the system, and then performed a qualitative case study that combined analyses of log data and the interactions among the students to understand students' reasoning processes in synergistic learning of physics and CT concepts.

## Collaborative learning and problem-solving

Roschelle and Teasley (1995) defined collaboration as "a coordinated, synchronous activity that is a result of a continuous attempt to construct and maintain a shared conception of a problem." Further research has shown the importance of developing shared understanding among the group members for successful task completion (Larkin, 2006). In addition, interaction skills of making and encouraging contributions of ideas, monitoring of progress, and providing constructive feedback through argumentation and explanation are essential components of collaborative learning (Garrison & Akyol, 2013; Grau & Whitebread, 2012). In the context of our work, this translates to students developing and utilizing their shared understanding of domain and CT concepts to co-construct their physics models, analyze and understand the behaviors generated by simulating their models, and where necessary, apply debugging processes to make refinements to and improve their models.

Working in close proximity and sharing a computer screen, provides students opportunities to explicitly discuss the model construction process, and develop arguments and explanations that support or challenge model constructs they propose or are proposed by their partners (Sins, et al., 2005). Therefore, students' interactive discourse provides opportunities for studying and understanding the reasoning processes they employ for synergistic learning during their computational model building activities. We briefly review discourse frameworks that have been developed for analyzing collaborative discourse in STEM domains.

## Collaboration discourse frameworks

The ICAP framework (Chi & Wylie, 2014) defines four different modes of engagement when considering learning behaviors: Interactive, Constructive, Active, and Passive. A learner who is engaging passively receives information but does not respond. An actively engaged learner receives information and manipulates it in some way. A learner who is receiving information, manipulating it and then constructing something with that knowledge is considered to be constructive. Finally, if two constructive learners are conversing about constructs they have generated from information, they are considered to be interactive.

The framework proposed by Weinberger & Fischer (2006) to analyze knowledge construction in a collaborative learning environment includes five categories of social modes: externalization, elicitation, quick consensus building, integration-oriented consensus building and conflict-oriented consensus building. Externalization is defined by a learner articulating their thoughts. Elicitation refers to the idea of using the other group partners as resources, this can be seen when a learner questions other group members. Quick consensus building is the act of accepting the contributions of another learner in order to continue progress. Integration-oriented consensus building occurs when one learner's understanding about a concept is changed based on another learner's reasoning. Conflict-oriented consensus building refers to the idea of different learners' perspectives contradicting each other, which leads to the debate and modification of conflicting ideas.

In our work, we superimpose Weinberger & Fischer's categories of social modes in knowledge construction on the ICAP framework to develop our coding scheme. This combination results in a more encompassing framework that considers a learner's social modes along with their mode of engagement on the collaboration task. A collaborative interaction is categorized by the highest level of engagement reached by one or more learners in the group combined with the type of social mode associated with the group. All three types of consensus building modes are only present when the collaboration is considered *interactive* because the group can only reach a consensus when both learners participate in model building. *Constructive externalization* or *elicitation* is categorized by a single student (the lead) is participating in model building and is narrating their actions and thought processes while the other students engage passively by following along silently, or actively, by verbally agreeing. During *constructive elicitation*, the lead student questions the other students in the course of building the model and receives little to no response. If questioning by two or more students is substantial and leads to responses that indicate the students are acting constructively, the collaboration is considered to be *interactive elicitation*. If neither student is participating in the model building but are reading aloud resources or instructions, the collaboration is categorized as *active externalization*. Finally, when none of the students are actively working on the model or saying anything they are in a *passive state*.

## Framing our research

### Learning physics by building computational models

We use C2STEM, a learning-by-modeling environment that incorporates a physics domain-specific modeling language (DSML) into a block-based programming environment to promote learning of domain-specific and CT concepts and practices through computational modeling and problem-solving exercises. C2STEM uses Netsblox (Broll, et al., 2017), an extension of Snap! (<http://snap.berkeley.edu/>). The use of a physics DSML aims to help students focus on physics concepts, while also helping students to write self-documenting code. In addition, we scaffold the model-building process by explicitly providing a simulation framework, where students can initialize variables and the use blocks to update the variables at each time step to capture dynamic behaviors.

Figure 1 illustrates a computational model that simulates the dropping of a package from a drone hovering at a specified height above a target. This requires students to think of the impact of gravity on the package's velocity and position with time that increments in steps of  $\Delta t$  when the simulation is run. The student also has to model a stopping condition, using a *conditional* construct to model the object's motion stopping when it hits the ground. Students can use graphing and tabulation tools shown as icons at the bottom of the stage.





Figure 1. Example final computational model for package drop.

## Synergistic learning

The notion of synergistic learning is predicated on the idea that the simultaneous learning of two domains in an integrated context can lead to better learning of concepts and practices in both domains than when the domains are learned separately. Previous work has shown that integration may initially increase the conceptual burden for students, but students are quickly able to overcome these difficulties and learn both domains better (Basu, et al., 2013; 2016). Our previous studies with C2STEM have shown that learning by building computational models of kinematics phenomena helps students gain a deeper understanding of the underlying kinematics concepts, while also helping them gain a better understanding of CT constructs, such as variable initialization, conditional, and loops (Hutchins, et al., 2018). Besides, a number of practices, such as abstraction, decomposition, and debugging transcend both domains (Sengupta, et al., 2013). However, to our knowledge, limited work has focused on identifying the synergistic STEM+CT processes that lead to these learning benefits.

In this paper, we identify some of the synergistic learning processes that we have observed in previous studies, and develop a framework for analyzing and understanding students' reasoning processes as they work through synergistic learning episodes during their model building tasks. In particular, we focus on the synergistic modeling practices of initialization, debugging, and conditional behavior changes. In our modeling framework, the initialization process can be identified by the addition, editing, and/or deletion of physics-DSML set blocks under the "When Green Flag Clicked" block (see Figure 1). Not only is variable initialization a key CT practice (Grover & Pea, 2013), students' selection of physics variables to initialize in the context of a particular task offers insight into their conceptual understanding of the motion processes being modeled. Debugging, a key CT skill (Grover & Pea, 2018), is related to analyzing model behaviors, identifying sources of error in the models, and then correcting the identified errors in physics modeling. In C2STEM, the debugging process is a truly synergistic process, as students must diagnose whether errors in their models result from incorrect representation (modeling) of physics concepts (e.g., an incorrect use of velocity in computing a look-ahead distance (specific point at which object slows down to stop) or an incorrect specification of the CT (or programming) constructs (e.g., writing the Boolean condition expression that initiates the slowing down behavior). Finally, we include the use of conditional expressions to indicate changes in motion behavior (e.g., slow down to avoid a collision). In all of these situations, the nature of the modeling tasks requires students to go back and forth between applying and checking their physics concepts and practices and their CT (or programming) concepts and practices to succeed in their model-building and model-checking tasks. We hypothesize that back and forth transition of concepts and practices across domains provides students with synergistic learning opportunities leading to better model building, and eventually better learning in both domains.

## Study description and data analysis methods

The qualitative research method presented in this paper is guided by the question: *What characteristics of synergistic learning and reasoning processes can we derive from students' collaborative discourse as they work on computational modeling tasks?* The primary data source used to answer this question was screen-capture video (using OBS™ software) that recorded students' actions in C2STEM, along with webcam video and audio, of two students ('S1' & 'S2') engaging in collaborative model building.

Our team conducted a two-month-long study that included 26 advanced sophomores assigned by our research team to work in groups of 2 or 3 on a kinematics and dynamics curriculum in C2STEM. The research team met with participants one school-day a week over a two-month period. Students completed one 45-minute CT training unit and four physics modules: three in Kinematics: 1D motion (with acceleration), 2D motion with constant velocity, and 2D motion with gravitational forces, and one in mechanics, i.e., an introductory unit on 1D Force. All students worked collaboratively, either in pairs or triads.

We used the rubric outlined in Table 1 to define and assess key learning objectives in physics and CT from the models that students constructed. The assessment results help us determine how successful students were in different aspects of model building and how their performance could be explained by their proficiency in synergistic learning and reasoning processes.

Table 1: Rubric for evaluating students Models

| Expressing physics relations in a computational model (physics component):  | Point(s) |
|---|----------|
| Program expresses correct relations among and units for needed Physics variables.   | 2        |
| Program reflects the effect of [velocity/acceleration] on [position/velocity] each time step.   | 1        |
| Program resulted in an accurate calculation of given task submission question.  | 1        |
| Using programming concepts to model physics phenomena (CT component)  | Point(s) |
| Program makes the distinction between actions that need to happen once during initialization and actions that need to be repeated in the simulation step  | 1        |
| Program correctly determines which actions always happen or happen under certain conditions   | 1        |
| Program updates the variable corresponding to the package's velocity: (1) under the correct conditions (e.g. correct conditional logic), and (2) in the correct fashion (e.g. each simulation step) | 2        |
| All code in the program is reachable and can be executed; No duplicate code   | 1        |

For analysis, we extracted and coded students' discourse mechanisms from the OBS video and voice recordings, capturing sequences of actions that that closely related to learning objectives in the two domains (see rubric), while also recording the challenges they faced, and how they overcame them. Two coders coded the above episodes. Inter-rater reliability was checked by calculating Cohen's kappa value which resulted in excellent agreement ( $k = 0.94$ ) for collaborative discourse and good agreement ( $k = 0.71$ ) for synergistic coding. For each task, we noted key actions and conversations that highlighted synergistic learning episodes and parsed these episodes to determine if the students' focus was on the physics or the CT aspect of their model.

### Student task performance in physics and CT

Our qualitative case study analyzed the model-building activities and the accompanying dialogue of two students who worked together on a laptop. We chose this group because they were the only pair in our study (all other students worked in triads), therefore, exchanges between them were easier to code. Besides this group was quite expressive, so we derived a lot of rich information from their dialog. Typically, one of them controlled the mouse and the keyboard, but the other student was always very attentive, and often initiated interactive dialog to discuss aspects of the modeling and debugging tasks. Their model scores (Figure 2) were assessed using the rubric described above on four model building tasks from each curriculum unit: 1D Motion, 2D Constant Velocity, 2D Motion with Gravity, Forces. As demonstrated, the group initially struggled in CT (3 out of 5), but improved to a perfect CT score by the final unit, even though the CT constructs and practices were more difficult. In physics, the group started off well, scoring a 3 out of 4 on the first two modeling tasks, improving to 4 out of 4 on the final modeling tasks. It is important to note that for both 1D motion and 2D constant velocity, the group did not receive a point for "program resulted in an accurate calculation of given task submission question," because this rubric item requires appropriate use of CT constructs. As such, further analysis of the modeling process is needed to understand issues related to transitions between physics and CT applications in the modeling process in order to understand the cause for their error.

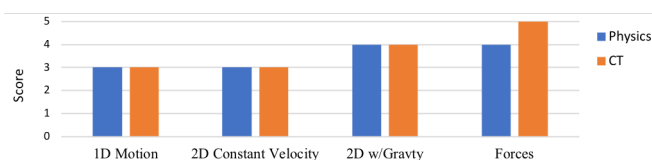


Figure 2. Group 1 model building scores.

### Case studies

Utilizing our synergistic learning framework) we have extracted segments of work and accompanying dialog that correspond to episodes of initialization, formulating conditional logic/stopping conditions, and debugging. The segments are presented in the order in which they occurred to study the students' progression in their model building skills and synergistic learning skills (Figure 2).

#### Segment 1: Conditional logic and stopping conditions

Episode 1 below describes a conversation and activity segments for a dyad (students S1 and S2) working on the 1D Motion module. In this task, students model the motion of a truck that speeds up from rest to a given

maximum speed (defined by a speed limit), maintain that speed and slow down and stop at a stop sign. This requires the students to calculate a lookahead distance from a stop sign. These two segments demonstrate their application of synergistic learning processes (their back and forth reasoning between the physics and CT concepts) to model the motion changes of the truck. We use the characterizations developed from the two collaborative frameworks to analyze and interpret the dialog constructs.

Table 2: Episode 1 (use of conditionals)

| Student's Words and Actions  | Physics and CT  | Collaboration   |
|--|---|---|
| <p><b>S1:</b> "15 m/s. So it needs to hit 15 m/s and then stop accelerating. So... If statements. If statements is the easiest way to do this."<br/> <b>S2:</b> "I thought you would do a when block."<br/> <b>S1:</b> "Why would you do a when block?"<br/> <b>S2:</b> "when the speed reaches..."<br/> <b>S1:</b> "Well, that's the thing because this starts a sequence, we need to put it inside the simulation step so that it will constantly repeat."</p> | <p><b>CT focused:</b> S1 and S2 agree on the physics concept but disagree on how to model it computationally, i.e., what CT construct(s) to use. S1 and S2 attempt to develop a shared understanding of different conditional operators</p> | <p><b>Interactive conflict-oriented consensus building:</b> S2 challenges S1's reasoning about a CT construct. S2 suggests a different idea and S1 pushes S2 to verbalize his reasoning.</p>  |
| <p><b>S2:</b> "Oh, I see."<br/> <b>S1:</b> "Right so if x velocity is greater than or equal to 15 then just change the x position by velocity. Else, change x velocity by acceleration."</p>   | <p><b>Physics focused:</b> Since S2 seems to agree, S1 brings the conversation back in context of physics. Using the language of the conditional block, S1 describes the relationships between position, velocity and acceleration.</p>     | <p><b>Interactive quick-consensus building:</b> S2 seems to agree with S1's reasoning but as seen in the continuation of the conversation, he is not fully convinced.</p>   |
| <p><b>S2:</b> "And you're telling me that's the easiest way?"<br/> <b>S1:</b> "That's the easiest way to do it because otherwise we have to do this and that's not a loop."</p>  | <p><b>CT focused:</b> S2 turns the conversation back to CT and challenges S1 again on the choice of conditional structure.</p>  | <p><b>Interactive conflict-oriented consensus building:</b> S2 challenges S1 again on the choice of conditional structure. After showing S2 on the screen, S1 and S2 develop a common understanding of the physics and CT concepts to use in their model.</p> |
| <p><b>S2:</b> "I would think like just like if velocity equals ... like if velocity equals 15 m/s set acceleration to 0 m/s..."</p>  | <p><b>Physics focused:</b> S2 attempts to support his reasoning by bringing in the relationship between velocity and acceleration.</p>  |   |
| <p><b>S1:</b> "We could do that but that would be... eh... I just I don't like the way that sounds cause yeah but yeah I know what you're saying."<br/> <b>S1:</b> "Ok so basically if velocity is equal...is greater than or equal to whatever. then change. then both of these...else just the bottom part. Ok."<br/> <b>S2:</b> "Oh I see why you put that there."<br/> <b>S1:</b> "Exactly"</p>  | <p><b>CT focused:</b> S1 shows S2 on the model how his idea would work</p>  |   |

In this segment, the dyad was having trouble converting their physics understanding into the computational constructs because they were unsure about the conditional constructs. Their consensus-building collaborative dialog demonstrates how they applied explanations, and argumentation to develop a shared understanding of the model building task (e.g., why select the "if" block and not a "when" block). However, their explanations are not deep, therefore, their justification for their model building steps is shallow. For example, S1 verbalizes his actions to develop a consensus and shared understanding with his partner, but the explanation "just doesn't like the way that sounds" implies an incomplete understanding of CT. In contrast, the physics knowledge is strong (including a correct calculation of the lookahead distance), and this is reflected in their task performance score.

### Segment 2: Initialization

The second episode in our case study involves the 2D constant velocity task focused on the initialization process. In this task, students needed to program a boat to cross a river, stopping at two islands located at different points on the way. The key physics concepts students have to learn are 2-D velocity, and how to compute the resultant velocity, given that the river current. In this episode, students are considering the

importance of initializing the heading variable of the boat given the need to change the boat's direction when moving towards a new target.

Table 2: Episode 2

| Student's Words and Actions  | Physics and CT  | Collaboration  |
|--|---|--|
| <p><b>S1:</b> "So it's 5 m/s in the x and y. So, we could set a different x velocity and different y velocity. Because it needs to go 15, so we could set the x to 5 and in the same three seconds if we set the y velocity to 2, then it would go 6 forward and 15 to the right."<br/> <b>S2:</b> (agreement sound)</p>   | <p><b>Physics focused:</b> S1 is verbalizing thought process on the relationship between x and y velocities and the respective distances moved in 3s</p>                | <p><b>Constructive externalization:</b> S1 externalizing his modeling constructs with S2 following along, occasionally agreeing</p>  |
| <p><b>S1:</b> "let's move this back and set the heading."<br/> <b>[S1 ADDS "set heading to" block and hardcodes it to the value 291.28. Clicks on the block to change direction of the boat and then removes block from stage.]</b><br/> <b>S1:</b> "so we need to set position to 0, -10.6"<br/> <b>[S1 ADDS and EDITS set position block]</b></p>  | <p><b>CT focused:</b> S1 adds an initialization blocks that supports his verbalization</p>  |  |
| <p><b>S2:</b> "How come you threw that block away?"<br/> <b>S1:</b> "What, that block? (pointing) Because we've already set the heading."<br/> <b>S2:</b> "Alright, but when you reset it's..."<br/> <b>S1:</b> "Right."<br/> <b>[S1 ADDS set heading block under GF and hardcodes to 291.28]</b><br/> <b>S2: jokingly says other student's name</b><br/> <b>S1:</b> "My wits have taken leave."</p> | <p><b>CT focused:</b> S2 challenges S1's removal of one of the blocks — presumably place in the wrong location causing the simulation to reset to an initial value.</p> | <p><b>Interactive conflict-oriented consensus building:</b> S2 challenges S1's action of discarding a set block, S1 tries to explain his reasoning and after further prompting by S2, sees the error</p> |
| <p><b>S1:</b> "And then, set position, set velocity"<br/> <b>[S1 ADDS set x velocity block under GF]</b></p>   |   | <p><b>Constructive externalization:</b> S1 reverts to narrating actions with S2 following along</p>  |
| <p><b>S1:</b> "and that's all we need to know, because it won't let us accelerate. It will let us accelerate in 2D air because that is when we start factoring in gravity. So then, start simulation, simulation step."<br/> <b>[S1 ADDS start simulation to GF and simulation step flag]</b></p>  | <p><b>Physics focused:</b> S1 concludes that they have completed the physics required for the model</p>   |  |

This segment illustrates a synergistic process where an understanding of the physics variables is needed to accurately model the object behaviors. We analyze how the students collaborate to resolve issues in the link between initialization and the modeling of the updates to capture dynamic behavior. The conflict-oriented consensus building approach allowed both students to consider the initialization process in the context of a complete model. By questioning the deletion, the group went from a predominantly constructive externalizing approach focused on the physics content, to an approach centered on generating consensus in understanding how the computational model should be set up. This conflict-oriented approach also occurred in Segment 1 in which the students also came to a consensus via questioning of the selection of a conditional block, but in this initialization scenario, we are beginning to see better justification as part of their reasoning. Although the improvements are observed in the synergistic discourse, the students' model score on this task was similar to Segment 1.

### Segment 3: Debugging

The final episode is an example of synergistic learning during debugging in a 2D gravity drop motion task, where the students modeled the delivery of two packages by a drone moving horizontally, calculating the look-ahead distance needed to release each package in order to safely land each at the desired targets on the ground. It is important to note that the group's score in both Physics and CT improved in this scenario, with the group developing a model that they used to correctly answer the task submission question.

Table 3: Episode 3

| Student's Words and Actions   | Physics and CT   | Collaboration  |
|---|--|--|
| <p><b>S1:</b> "Did we miscalculate? Did we miscalculate? Does it need to be like 9 meters or something? Let's try 9 meters just to be sure. I have a sneaky, sneaky suspicion."</p>   | <p><b>Physics focused:</b> S1 is pointing out that the physics calculations put into the model may be incorrect.</p>   | <p><b>Interactive integration-oriented consensus building:</b> S1 and S2 work together to find the error in their model, and they conclude that it is likely a time miscalculation</p> |
| <p>[S1 edits subtraction in if via hardcoded].<br/> <b>S1:</b> "Let's try this again." [S1 presses play]<br/> <b>S1:</b> "Drop..."<br/> <b>S2:</b> "It's not..."<br/> <b>S1:</b> "Yeah, that's not right."<br/> <b>S2:</b> "Wait, I want to see it" [S2 takes control of mouse]</p> | <p><b>CT focused:</b> S1 and S2 are using the model to determine why the package is not ending up in the correct place</p>   |  |
| <p><b>S1:</b> (inaudible) "Did we miscalculate the time?"<br/> <b>S2:</b> "We might have"<br/> <b>S1:</b> "We might have miscalculated the time. Let's go back and look at the time equation. We could do this one, too, couldn't we?"<br/> <b>S2:</b> (agreement sound)</p>        | <p><b>Physics focused:</b> After their model does not work as expected, S1 and S2 go back to determining where they made an error in modeling the physics relations.</p> |  |

This segment illustrates how the students use the animation of the object's motion to realize they have made an error, and then work together to find its source. The discussion is mostly physics-focused, but it does require them to analyze how the sequence of blocks they have used to build the model relates to the behavior they are observing. Their suspicion is that they did not model the release time of the object correctly. The episode does show some switching between a physics and CT focus in their conversations. This was also observed in the earlier episodes, but this one demonstrates a level of maturity in that they are not arguing over what construct to use, but are trying to related behavior back to model structure, which implies higher level synergistic reasoning.

## Discussion

When the S1 and S2 are successful in their modeling tasks, the conversation has the characteristic of integrating both CT and physics reasoning, and adopts a back and forth between the two domains in the model construction task. We believe that this back and forth, i.e., analyzing relevant domain concepts required for modeling and representing them using computational constructs, and going back and forth to establish the correctness of the physics and its computational representation is a key element of synergistic learning and reasoning. A conversation that integrates both CT and physics reasoning does not necessarily imply students will succeed, as seen in Episode 3. The students go back and forth between the physics concepts and the CT construct in their conversation, but run out of time before they come up with the correct form of the model. However, they gain a synergistic understanding that they apply later in their modelling process. The dyad also discusses look-ahead distance calculations in the 1D constant velocity module, but this earlier conversation does not have them switching between physics and CT reasoning. The discussion that can be split into two completely separate conversations; one where students determine the physics calculations separately and then a second discussion about the computational modeling. When comparing the progress made between episode 1 and 3, it is clear that the students are better able to integrate their physics and CT knowledge. Their synergistic learning gains are shown as their modeling score increase in physics and CT from the 1D constant velocity to Forces modules (Figure 2).

## Conclusions and future work

One challenge of using video analysis to study synergistic learning is recognizing evidence of synergistic learning in practice. Another challenge in using video analysis to study synergistic learning is knowing where to look, amongst the many hours of video data. We have found evidence of synergistic learning in three specific contexts: when students were (1) using conditional blocks, (2) initializing variables, and (3) debugging code. These findings will help us hone our search for episodes of synergistic learning in future research.

Focusing on one student dyad, we first found evidence of synergistic learning, initially in their inability to reconcile the physics conceptual knowledge with the computational constructs required to construct the computational model. This involved a lot of dialog about whether the physics knowledge was correct, and then whether the computational construct reflected that the physics knowledge had been applied correctly. Studying the students' conversations in this phase provides us with some understanding of their difficulties, which can be

in the domain concepts, the computational constructs, or the representation of the domain concepts correctly in the computational form. It is also clear from the students' dialog that such exercises force them to think deeper about both the domain concepts and the computational constructs. Initially, this may increase their difficulties. However, by executing their computational representations, i.e., simulating their model, they have the opportunity to implement debugging processes that may help them understand and overcome their difficulties. It also provides us with opportunities to detect such episodes, and adaptively scaffold students who are unable to overcome their difficulties. In our particular case study, the two students succeeded in working through their difficulties on their own, after some initial stumbles.

Our next steps include expanding our analysis to a broader group of students and extending our analysis to gain insights into collaborative regulation.

## References

- Basu, S., Dickes, A., Kinnebrew, J. S., Sengupta, P., & Biswas, G. (2013). CTSiM: A Computational Thinking Environment for Learning Science through Simulation and Modeling. In *CSEDU* (pp. 369-378).
- Basu, S., Biswas, G. & Kinnebrew, J.S. (2016). Using multiple representations to simultaneously learn computational thinking and middle school science. In Thirtieth AAAI conference on Artificial Intelligence (pp. 3705-3711). Phoenix, Arizona, USA.
- Basu, S., Biswas, G., & Kinnebrew, J.(2017). Learner modeling for adaptive scaffolding in a computational thinking-based science learning environment. *User Modeling & User-Adapted Interaction*, 27(1), 5-53.
- Broll, B., Lédeczi, A., Volgyesi, P., Sallai, J., Maroti, M., Carrillo, A., Weeden-Wright, S., Vanags, C., Swarts, J., & Lu, M. (2017). A visual programming environment for learning distributed programming. In *Proceedings of SIGCSE 2017*. ACM.
- Chi, M. T., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, 49(4), 219-243.
- DiSessa, A. A. (2001). *Changing minds: Computers, learning, and literacy*. MIT Press.
- Garrison, D. R., & Akyol, Z. (2013). The Community of Inquiry Theoretical Framework. In *Handbook of distance education* (pp. 122-138). Routledge.
- Grau, V., & Whitebread, D. (2012). Self and social regulation of learning during collaborative activities in the classroom: The interplay of individual and group cognition. *Learning and Instruction*, 22(6), 401-412.
- Grover, S., & Pea, R. (2013). Computational thinking in K-12: A review of the state of the field. *Educational Researcher*, 42(1), 38-43.
- Grover, S. & Pea, R. (2018). Computational Thinking: A competency whose time has come. In *Computer Science Education: Perspectives on teaching & learning*, Sentance, S., Barendsen, E. & Carsten, S. (Eds). Bloomsbury.
- Hutchins, N., Biswas, G., Maroti, M., Ledezci, A., & Broll, B. (2018). A Design-Based Approach to a Classroom-Centered OELE. In Intl. Conf. on Artificial Intelligence in Education (pp. 155-159). Springer.
- Larkin, S. (2006). Collaborative group work and individual development of metacognition in the early years. *Research in science education*, 36(1-2), 7-27.
- Roschelle, J., & Teasley, S. D. (1995). The construction of shared knowledge in collaborative problem solving. In *Computer supported collaborative learning* (pp. 69-97). Springer, Berlin, Heidelberg.
- Sengupta, P., Kinnebrew, J. S., Basu, S., Biswas, G., & Clark, D. (2013). Integrating computational thinking with K-12 science education using agent-based computation: A theoretical framework. *Education and Information Technologies*, 18(2), 351-380.
- Sins, P., Savelsbergh, E., & van Joolingen, W. (2005). The Difficult Process of Scientific Modelling: An analysis of novices' reasoning during computer-based modelling. *Intl. Journal of Sci. Ed*, 27(14), 1695-1721.
- Weinberger, A., & Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & education*, 46(1), 71-95.
- Wilensky, U., & Reisman, K. (2006). Thinking like a wolf, a sheep, or a firefly: Learning biology through constructing and testing computational theories—an embodied modeling approach. *Cognition and instruction*, 24(2), 171-209.
- Wing, J. (2006). Computational thinking. *Communications of the ACM*, 49(3), 33-35.

## Acknowledgments

We thank our C2STEM collaborators at Vanderbilt University, SRI International, Stanford University, and ETS. This research is supported by NSF grant #1640199.

# Girls as Experts, Helpers, Organizers, and Leaders: Designing for Equitable Access and Participation in CSCL Environments

Kay E., Ramey, Northwestern University, kay.ramey@northwestern.edu  
Reed Stevens, Northwestern University, reed-stevens@northwestern.edu

**Abstract:** Within the CSCL community and in computing and computational making more broadly, issues of equity continue to be under-researched and undertheorized. Here, we examine how FUSE Studios – a set of in-school, choice-based, STEAM learning environments, based around a set of digital and tangible making challenges – supports equitable access to and participation in making and computing. Drawing on web-log data and video-ethnographic data, we argue that four characteristics of FUSE support equity: the design of the challenges; the diverse ways of knowing and doing supported by the activity system; the specific interactions encouraged by the activity system; and the program’s placement inside school. We focus, here, on gender equity, but also discuss implications and planned research on other aspects of equity.

## Introduction

Within the CSCL community, issues of equity continue to be under-researched and undertheorized. Of the articles published in the *International Journal of Computer-Supported Collaborative Learning* since it started in 2006, only one title references anything related to women or gender. Further, despite the theme of the 2017 conference, “Making a Difference: Prioritizing Equity and Access in CSCL”, of the over 200 contributions to the conference proceedings, only six listed “equity and gender” as a keyword. Only 61 mentioned either “gender” or “equity” *at all* in the text, and far fewer dealt with either issue substantively, with many mentions of “gender” appearing only in a description of participants or groupings and many mentions of “equity” appearing only as authors quoted the conference theme. This lack of attention to equity is reflective of larger inequities in participation in computing and computational making (e.g., Funk & Parker, 2018; Lewis, 2015; Margolis & Fisher, 2002; Vossoughi & Bevan, 2014). Despite well-documented equity problems in these fields, efforts to engage women and minorities are still far too few and have been only minimally successful at involving a diverse audience in equitable ways.

## Background

Research that has addressed equity in computing and computational making has done so in two ways. The first is by attempting to create *equitable access* to these activities and spaces by removing barriers to entry for women, minorities, and less affluent individuals (e.g., AAUW, 2000; Lewis, 2015). For example, Lewis (2015) found that many women avoid makerspaces, because they don't have prior computing experience or struggle with the absence of clear goals in making activities. However, multiple studies suggest that if we can get females in the door and get them engaged, they are more likely to express increased confidence and empowerment related to the tools and skills involved (e.g., Barniskis, 2014; Bowler, 2014; Fields & King, 2014). As a result, some recent efforts have focused on bringing making and computing into schools, where there is greater access (e.g., Atit et al., in press).

However, efforts to diversify access to making or computing have not necessarily resulted in *equitable participation* (e.g., Scott, et al., 2015; Vossoughi & Bevan, 2014; Vossoughi et al., 2013). Thus, other studies (e.g., Vossoughi et al., 2013) have focused on understanding what types of knowledge and interactions are valued in a learning space and used this as a lens to see how the space does or doesn't create opportunities for young people of different backgrounds to experience themselves as knowledgeable contributors. For example, Lachney, et al. (2016) argued that the technologies that designers claim are value-free may actually be implicitly exclusionary. Vossoughi and Bevan (2014) also argued that makerspaces have the potential to challenge deficit views and support more inclusive learning *if* they support interactions such as novices and experts working side-by-side, assisting each other, and continually shifting roles. However, *if not*, making and computing activities also have the potential to reproduce existing inequities. For example, Volman and van Eck (2001) documented a number of studies of learning through computer games where boys took over the computer, sidelining girls.

A handful of approaches have recently emerged that seek to engage females and underrepresented minorities in making and computing. One successful example of this is e-textiles (e.g., Barniskis, 2014; Buchholz, et al., 2014; Buechley, et al., 2008; Fields & King, 2014; Kafai, Fields & Searle, 2014). This approach integrates craft skills (textile work) with electronic circuitry and computer programming. Researchers have found that females prefer to learn new technology skills by incorporating them into such craft work (Barniskis, 2014; Kafai, et al., 2014). Others have found that e-textiles disrupt preconceptions about gendered ability, access, and authority in making (Buchholz, et al., 2014). Still others have used e-textiles to create cultural connections between Native American indigenous practices and computational thinking (Kafai, Searle, Martinez & Brayboy, 2014). However,

the specificity of this approach invites the questions, ‘What about students who aren’t interested in textiles?’ and ‘What about skills that can’t be taught through this (or any) single activity?’ This tool-centered approach also diverts focus away from the role that activity systems in which tools are embedded play in promoting equity.

Another intervention that focused more holistically on providing supportive infrastructure for equitable computational learning was “Digital Youth Divas” (DYD; Pinkard, et al., 2017). This out-of-school program used narrative stories to motivate the creation of digital artifacts and support non-dominant, middle-school girls’ STEM interests and identities. In addition to the narrative-driven, project-based curriculum, DYD involved an online social network, supportive adult mentors, and support for the extension of interests beyond the DYD context.

Here, we explore a program, FUSE Studios (e.g., Stevens et al., 2016), which shares DYD’s attention to the role of sociomaterial infrastructure in supporting equitable interest development and learning but takes a different approach to achieving these goals. FUSE is a set of in-school, choice-based, STEAM learning experiences, structured around a set of digital and tangible making challenges. Drawing on web-log data, video-ethnographic data, and interviews with students and teachers, we will show how FUSE supports equitable access to and participation in making. We will also show how it supports learning that is embodied, enactive, extended, and embedded, the conceptual themes of this year’s conference. Elsewhere, we have written about the affordances of this activity system for supporting: (1) a variety of collaborative learning arrangements (Penney, 2016; Stevens et al., 2016); (2) the development and sharing of expertise (Penney, 2016; Stevens et al., 2016); (3) interest development (Ramey, 2017; Ramey & Stevens, submitted); and (4) learning (Ramey, 2017; Ramey, Stevens, & Uttal, submitted). Here, we examine whether and how these outcomes were equitably available. We focus the discussion of equity here on gender equity, because this is where we have the most complete data. However, in the Conclusion section, we also discuss implications for and future research about other dimensions of equity.

## Methods

### Research context

FUSE is structured around a set of almost 30 STEAM challenge sequences, housed on the FUSE website (fusestudio.net). These challenges level up like video games, with each challenge containing multiple levels of increasing difficulty and complexity. The challenges are designed to build on student interests (e.g., video games, jewelry or clothing design, music), as well as make connections to the tools and activities of STEAM professionals (e.g., CAD, 3D printing, programming, electronics). FUSE is structured to allow students choice in what challenges to do, how to approach challenges, and what resources (e.g., other students, physical and digital materials) to draw upon for help. In other words, FUSE is more choice-based, student-driven, and interdisciplinary than traditional learning in schools—or even learning in workshop-style making activities, where all students do the same project or use the same tools (e.g., Fields & King, 2014). However, it also provides more structure and support than many makerspaces, where learners are let loose to explore available tools without being provided clear goals (e.g., Brahms, 2014). Further, although FUSE started as an after-school program, it is now primarily being implemented in schools, as a standalone class. We argue that the following four characteristics of FUSE – (1) the design of the challenges; (2) the diverse ways of working and learning supported by the activity system; (3) the specific interactions encouraged by the activity system; and (4) the program’s placement inside the school day—promote both equitable access to and equitable participation in this learning environment for girls and boys.

### Data collection and analysis

The first way in which we examined gender equity in FUSE was to examine challenge activity data from the FUSE website. Every student who creates an account on the website is asked to indicate their gender. Once users set up accounts, they are met with a gallery of challenges to choose from. They are also able to watch trailer videos to invite interest in particular challenges. Once they identify a challenge of interest, they can choose to “start” level one of the challenge (an action that the website logs). In order to unlock subsequent levels (up to five), users must “complete” the previous level by uploading an image, video, or digital artifact that demonstrates they have completed that level (another action that the website logs). For our analyses here, we drew on these three data points (gender, level starts, and level completes) to look for equity in challenge interest and persistence.

The second way in which we analyzed equity in FUSE was by drawing on data from video-ethnographic observations of seven FUSE classrooms, from one large, suburban, racially- and socioeconomically-diverse school district, over the course of the 2015-16 school year. Each of these classes met twice per week for a total of 90 minutes, and a member of our team was present for all sessions, acting as a participant observer. During each visit, we took field notes and recorded whole-room video using a stationary camera and point-of-view video using visor-mounted cameras worn by six focal participants per class. At the end of the school year, we also conducted semi-structured interviews (Patton, 1990) with 60 focal students and seven facilitators to understand what they



had learned or remembered from FUSE and what impact, if any, it had had on their interests or identities.

In analyzing this ethnographic data, we drew on interaction analysis (e.g., Goodwin, 2000; Hall & Stevens, 2015; Jordan & Henderson, 1995), because, in line with the conference theme, it allowed us to analyze ways in which knowledge was embodied, enactive, and embedded in this particular sociomaterial context. We also selected this method based upon work by Vossoughi et al. (2013) and DiGiacomo & Gutiérrez (2016), which proposes that equity is produced and reproduced (and therefore can be analyzed) interactionally.

## Findings

### Evidence from web-log data of equitable access and participation



We analyzed data from the FUSE website in three ways to examine whether equitable access and participation were achieved for girls in FUSE. First, an examination of the gender of all users on the FUSE website showed that over the lifetime of the program (2012-2018), more girls have participated in the FUSE than boys (22068 versus 21126). This indicates that we have been successful at getting girls in the door, or achieving equitable access. Second, an analysis of persistence shows that boys and girls persisted at similar rates. For example, 39% of girls and 40% of boys who started level one of a challenge went on to start level two. Similarly, 56% of girls and 57% of boys who started level two of a challenge went on to start level three. Finally, an analysis of challenge interest by gender—where “interest” was defined as starting level 1—showed that of the 26 challenges available at the end of 2018, 14 were preferred by boys, 9 were preferred by girls, and 3 were gender neutral (see Table 1). This indicates that we’re doing relatively well developing challenges that appeal to both genders, but there is still room for improvement in developing more challenges that appeal to girls.

Table 1: Challenge preferences by gender

| Gender Neutral Challenges (3)  | Challenges Girls Prefer (9)   | Challenges Boys Prefer (14)   |
|--|---|---|
| Smart Castle<br>Dream Home 2: Gut Rehab<br>Eye Candy   | Jewelry Designer****<br>Just Bead It!****<br>Selfie Sticker****<br>Cookie Customizer****<br>Electric Apparel****<br>Spaghetti Structures****<br>Keychain Customizer****<br>Dream Home****<br>LED Color Lights** | Print My Ride****<br>Get in the Game****<br>Game Designer****<br>Laser Defender****<br>Solar Roller****<br>Music Amplifier****<br>Wind Commander****<br>Ringtones****<br>Coaster Boss****<br>Minime Animation****<br>How to Train Your Robot****<br>VR Escape Room****<br>Mini Jumbotron****<br>3D You*** |
| *p < .05, **p < .01, ***p < .001, ****p < .0001. Results are from a chi-square test of level 1 starts from a sample of 43194 users (22068 girls, 21126 boys) between the start of FUSE in July 2012 and December 2018. The more statistically significant a p value, the more ‘gendered’ a challenge is. |   |   |

The good news is that this web data analysis also helps us understand where specific improvements can be made. For example, from an analysis similar to the one presented above, conducted on data from 2012-2015, we noticed that our only robotics challenge (*Robot Obstacle Course*) skewed heavily male. So, we redesigned it to use a different robot with different appearance and a different programming language. We also changed the challenge name, trailer video, and goals to reference different cultural imagery (*How to Train Your Robot*). As a result, girls’ participation in the challenge increased by 12% (see Table 2).

Table 2: Redesigning a robotics challenge to make it more interesting to girls

|       | Robot Obstacle Course (ROC)   | How to Train Your Robot (HTYR)  | Change   |
|-------|---|---|--|
| Robot |  |  | +Pre-built<br>+References different cultural imagery (training a pet vs. a tactical mission) |

|                           |       |       |   |
|---------------------------|-------|-------|---|
|                           |       |       | +transferable programming               |
| % Level 1 Starts by Girls | 33.7% | 45.4% | 12% increase in level 1 starts by girls |

### Ethnographic evidence for equitable participation

Analysis of video-ethnographic data also provided evidence that girls participated equitably in learning and developing interests in FUSE. Here, we present three cases highlighting key features of FUSE that promoted equitable participation. These cases are representative of both the ways in which girls participated in FUSE and the positive experiences they reported having in their end-of-year interviews. For example, of the 32 girls we interviewed, none reported negative experiences in FUSE and 30 reported highly positive experiences.

#### Johanna, Victoria, and Andrea: Girls organizing their own learning

Our first case involves a group of fifth grade girls, Johanna, Victoria, and Andrea, who spent their entire school year in FUSE working together on challenges. They began with the *Dream Home* challenge, each creating their own home using the CAD program, Sketchup, but sitting together, helping each other, and showing each other things they'd made. After finishing all three levels of *Dream Home*, they worked through a follow-up challenge, *Dream Home 2: Gut Rehab* and a vinyl cutting challenge called *Selfie Sticker*, again making their own, individual products but sitting side-by-side and helping and consulting with each other. The episode in Table 3 shows how Johanna sought help from Andrea and Victoria when she encountered a problem in *Dream Home*. Here, she was trying to see inside her CAD model. Victoria had just figured this out moments before. When Andrea saw her make this discovery, she asked Victoria, "How are you doing that? How'd you get it to that view?", and Victoria showed her. So when Johanna asked Victoria the same question (line 1), Andrea was able to show Johanna.

Table 3: Andrea shows Johanna how to get "inside" her CAD model home

| Line | Person   | Talk   | Actions  |
|------|----------|--|--|
| 1    | Johanna: | Victoria, how do you get inside?               |  |
| 2    | Andrea:  | Got it! I got it, I got it, I got it!          |  |
| 3    | Johanna: | How do you? How do you do it? Can you tell me. |  |
| 4    | Andrea:  | Ok, so you're going to go to like the feet.    |  |
| 5    | Johanna: | I am at the feet.                              |  |
| 6    | Andrea:  | Uh huh.  |  |
| 7    | Johanna: | Mmm hmm  |  |
| 8    | Andrea:  | And then <sup>1</sup>                          | <sup>1</sup> Takes Johanna's mouse, scrolls forward. |
| 9    | Johanna: | And just zoom in or?                           |  |
| 10   | Andrea:  |  | Scrolls forward until inside is visible.             |

This interaction shows how the girls shared knowledge to accomplish challenge goals. For example, Johanna sought help from Andrea and Victoria (lines 1 and 3), and Andrea provided help by explaining, then showing Johanna how to use the tools in Sketchup to achieve her goal (lines 4, 8, and 10). It also shows how Johanna actively participated in her own learning (lines 5, 7, 9), rather than passively letting Andrea do things for her. Finally, the fact that what Andrea was showing Johanna here was a skill she had just learned from Victoria shows how fluidly the girls were able to shift roles from novice to expert. This interaction not only shows how learning in FUSE was enactive but also how knowledge was embodied. For example, by asking "Victoria, how do you *get* inside?" Johanna was drawing on the embodied metaphor of walking around inside of the virtual CAD world and physically entering her model. This metaphor was supported by the tools in Sketchup, such as "the feet" (lines 4 and 5), a tool with an icon that looks like footprints and allows users to "walk" through virtual space. Andrea also shared her knowledge with Johanna in an embodied way, taking her mouse (line 8) and showing her, through embodied action, how to manipulate the hardware and software tools to achieve her goal.

As a result of many interactions like this one, not only were all three girls able to proceed through challenge levels of increasing complexity and master a set of technical tools and skills, but they also developed STEAM interests and identities related to their challenge work. For example, by her end-of-year interview, Johanna explained why she hoped she would be able to do FUSE again the next year, by saying, "...*this kind of helped me decide that I wanted to be an architect*, after my mom said it, because it's fun to make your own things."

We argue that Johanna, Victoria, and Andrea's collaborative learning arrangement and the learning and interest development that stemmed from it were possible in FUSE, because they were allowed to choose who to

work with, what to work on, and how to approach the work. In other words, they were able to organize their own learning in ways that were productive for them. Their all-girl learning arrangement represents a significant departure from male-dominated computing and making spaces, in which not only is the culture dictated by males, but there may be so few females that girls are forced to seek help from males to achieve goals and/or be sidelined while boys take over control (Volman & van Eck, 2001). The fact that the girls were able to choose what to work on also allowed them to choose challenges which are broadly favored by girls but which are still technically challenging. Finally, the fact that they could work on challenges with prescribed goals allowed them to overcome one of the barriers to entry for females in makerspaces – the lack of clear goals (Lewis, 2015)

### Erin: Girls leading others' learning

Our second case is Erin. Like Johanna, Victoria, and Andrea, Erin worked on challenges with other students. However, unlike Johanna, Victoria, and Andrea, she worked with a variety of, mostly male, students. Unlike girls making their way into male-dominated spaces, in Erin's interactions with boys during FUSE, she was consistently the leader. The episode in Table 4 is from Erin's work with two boys, Ajay and Aiden, as they did the *Solar Roller* challenge together. The goal of this challenge is to build a solar car capable of travelling a fixed distance along a track. In this interaction, the students were trying to add a capacitor to their car so that it could travel through a 50-inch tunnel. In this interaction, Erin directed her group's work (line 1, 12, 14), acquired problem-solving resources (lines 2, 10, 14, 16), and explained to Ajay what a capacitor was (line 18). Meanwhile, the boys messed around (lines 3-8), contributed minimally (lines 9, 11, 13, 15) and deferred to Erin's expertise (line 17).

Table 4: Erin leads her group in wiring a capacitor into their solar car

| Line | Person | Talk   | Actions  |
|------|--------|--|--|
| 1    | Erin:  | <sup>1</sup> Okay, okay, Aiden, we're going to have to set up the capacitors.  | <sup>1</sup> <i>Goes back and looks at the directions.</i>   |
| 2    | Erin:  | Where's the bread board? <sup>1</sup> Oh here it is. <sup>2</sup> Which one? The big one. The big one. This one's the positive side.   | <sup>1</sup> <i>Looks in supply box.</i> <sup>2</sup> <i>Looks at directions and diagram.</i>  |
| 3    | Boys:  |  | <i>Laughing.</i>   |
| 4    | Erin:  | What? <sup>1</sup>   | <sup>1</sup> <i>Turns toward boys.</i>   |
| 5    | Ajay:  |  | <i>Laughing.</i>   |
| 6    | Erin:  | What?  |  |
| 7    | Ajay:  | He put his finger on the super glue thing. <sup>1</sup>  | <sup>1</sup> <i>Holds up a plastic bag.</i>  |
| 8    | Aiden: | I just picked it up, and it just gacked glue I guess.  |  |
| 9    | Ajay:  |  | <sup>1</sup> <i>Rummages through box.</i>  |
| 10   | Erin:  | Oh, I already got all the stuff. <sup>1</sup> Alrighty, so we need to...so the solar panel's right here. <sup>2</sup> The motor's right here. <sup>3</sup>   | <sup>1</sup> <i>Looks at directions.</i> <sup>2</sup> <i>Puts panel in place.</i> <sup>3</sup> <i>Puts motor in place.</i>                               |
| 11   | Ajay:  | Here, first...   |  |
| 12   | Erin:  | No, stop stop stop stop!   |  |
| 13   | Ajay:  | Start with the car. Then see what you can do with it.  |  |
| 14   | Erin:  | No, don't:::oh then that, yeah. <sup>1</sup> Ok, so then we're going to have to put the negative side in one of these things <sup>2</sup> and the positive side <sup>3</sup> :::where the positive side:::and don't ask why I know which one's negative. | <sup>1</sup> <i>Looks at diagram.</i><br><sup>2</sup> <i>Inserts negative leg of capacitor into breadboard.</i><br><sup>3</sup> <i>Looks at diagram.</i> |
| 15   | Ajay:  | We should do like more research at our houses.   |  |
| 16   | Erin:  | If I could fit this thing in here. <sup>1</sup> There I go. I'm just gonna add it. Ok so this is the capacitor, and it would, <sup>2</sup> short leg on the capacitor. Move the setup on the bread board. Wait what?                                     | <sup>1</sup> <i>Inserts positive leg of capacitor into breadboard.</i><br><sup>2</sup> <i>Reads directions aloud.</i>                                    |
| 17   | Ajay:  | What's a capacitor?  |  |
| 18   | Erin:  | Um, it like gives energy, a short burst of energy, once the light disappears, continues the loop, but I don't get what <sup>1</sup> this is.   | <sup>1</sup> <i>Points to diagram on screen.</i>   |

In her role as group leader, Erin correctly installed the capacitor and came to an embodied understanding

of electrical circuits and the role of capacitors in them. We see this in line 18, when Erin provided a definition of a capacitor that incorporated the spatial arrangement of the circuit she had just constructed and the capacitor's role in it ("continues the loop"). By providing this explanation, Erin showed how knowledge is enacted through embodied activity embedded in a particular sociomaterial context. Erin also showed how, in FUSE, girls were able to be leaders, rather than followers, in learning. Like Johanna, in Erin's end-of-year interview, she reported emerging STEAM interests related to FUSE. When asked "Have you thought about what you want to do when you grow up?", Erin responded, "Yeah, I want to, since I really like space and stuff like that, I would like to be an astrophysicist or chemist...I would like to also be a programmer or stuff like that." This is significant, because in addition to doing *Solar Roller* and a number of CAD challenges, Erin spent substantial time working on computer programming challenges, such as *Game Designer* (released midway through the 2015-16 school year).

In our end-of-year interview with Erin's teacher, she also emphasized Erin's leadership role, saying, "Erin also started out with the dream house, but she is so into the computers and into GEMS and that, that *she actually took a lot of her activities home...Her group*, when they did the activity with the cars to get them to move, it was just a lot of *fun to problem solve with them*." This quote suggests that Erin's teacher recognized her role as the group leader in the *Solar Roller* challenge ("her group"). It also shows that she recognized the extended nature of Erin's interest development and learning in FUSE, as she mentioned Erin working on activities at home (without being assigned to) and participating in extracurricular STEAM activities (GEMS – Girls Excelling in Math and Science). This quote also suggests of how this teacher saw her role in FUSE – as a facilitator or helper rather than instructor (e.g., "fun to problem solve *with them*"). This indicates one mechanism through which space was made for Erin or other students to step into leadership roles – because the teacher was not dominating them.

#### Carmen: Girls' expertise being valued

Our final case is a student, Carmen, who represents of a type of interest, identity development, and learning frequently observed in FUSE—students developing interests and *relative expertise* (Stevens et al., 2016) related to 3D printing, using that expertise to help others print, and becoming recognized as *relative experts*. For example, the transcript in Table 5 shows Carmen helping another student, Elena, print by fixing a problem with the printer.

Table 5: Carmen fixes a problem with the 3D printer and manages the print queue

| Line | Person      | Talk   | Actions  |
|------|-------------|--|--|
| 1    | Carmen:     | Something must have went wrong with um, when it was pulling this in, <sup>1</sup> it must have gotten a little bit tangled.                    | <sup>1</sup> <i>Pulls on guide tube surrounding filament.</i>  |
| 2    | Elena:      | <sup>1</sup> Can I edit this? <sup>2</sup>   | <sup>1</sup> <i>Sits down at computer connected to printer.</i> <sup>2</sup> <i>Looks at Carmen.</i>         |
| 3    | Carmen:     | What? No you can't, Elena. Sorry. Unless you want to do it on this computer, and then let Diego print today, and then you can print next week. |  |
| 4    | Elena:      | I'm not going to print next week <sup>1</sup> , unless I'm going, unless I go after Diego?   | <sup>1</sup> <i>Shakes head.</i>   |
| 5    | Carmen:     | You can go after Diego <sup>1</sup> , right after Diego.   |  |
| 6    | Aaliyah:    | This is probably going to take a little bit longer.  |  |
| 7    | Carmen:     | So do you want to print after Diego or do you want to print now, with 'Focus'? Your choice.  |  |
| 8    | Aaliyah:    | Just make it really big  |  |
| 9    | Researcher: | Is there something that you could do to fix it, so that it will print better next time, do you think?  |  |
| 10   | Elena:      | Maybe like, I want to make it  |  |
| 11   | Carmen:     | Something, it probably got tangled right here <sup>1</sup> in the string and it wasn't going in through it <sup>2</sup> .                      | <sup>1</sup> <i>Reaches for filament, then points to it.</i> <sup>2</sup> <i>Points to extrusion nozzle.</i> |

In lines 1 and 11 of the transcript, Carmen demonstrated her embodied technical knowledge and troubleshooting skills with the printer by presenting a possible solution to a problem both in words and embodied action. She then fixed the problem by untangling the filament. Elena positioned Carmen as an expert by seeking her guidance (line 2), letting her solve the problem, and deferring to her authority about the print queue (indicated by Elena's lack of response after line 5). Two weeks later, another student, Diego, also recognized Carmen's expertise, by awarding her the class's travelling "Engineer Award". He did so in recognition of both her technical expertise at

3D printing *and* her use of that expertise to help other students print, saying, “I nominate Carmen, because how she was working hard and how she was like helping us and like pushing herself and as soon as something was wrong, she would fix it.” Diego’s nomination was seconded by a number of students and the teacher.

Carmen’s teacher also showed recognition of her 3D printing expertise in his end-of-year interview, saying “Carmen, she became like our 3D printer guru. You know, she was the one that you would go to with any 3D printer issue. She could change it. She could do it”. He contrasted this with her confidence elsewhere, saying, “And you know, in class, she might not always show that confidence. Um, you know, she likes to participate, but you can tell with her answers sometimes, there’s not the confidence, whereas here at FUSE, when it came to 3D printing, Carmen could tell you with confidence.” This suggests that the forms of knowledge and interaction valued in FUSE allowed Carmen to build confidence and expertise in ways unavailable in a traditional classroom.

Carmen also recognized her own expertise. For example, one day, when the first author asked Carmen and Elena, “What are you ladies working on today?” Carmen replied, “Uh, actually, I’m helping her print, *because I’m like the master of the printer now*, and the computer.” In Carmen’s end-of-year interview, she explained her interest in the 3D printer, saying “I like working on the 3D printer, and I like helping other people, um, with the 3D printer so that they can print and they can have their prints and be happy with it.” In other words, her interest was both in the technical object but also in using her technical work with it to help others. Carmen also connected this work to her emerging identity and interests, saying “So I am kind of like a generous person helping...and when I grow up I wish to help cancer kids and become a doctor for them...So I’m starting now and helping people with the 3D printer...the 3D printer is like a cancer kid. I get to help it. If it’s broken, I get to cure it and fix it.” Carmen’s words show how FUSE supported her in discovering interests and identities, developing social and technical skills, and being recognized for those skills by her peers and teacher. In other words, they show how the FUSE culture valued her ways of knowing and allowed her to experience herself as a knowledgeable contributor.

## Conclusion

The analyses presented here show how CSCL environments can be designed, studied, and improved to promote equity. Using web-log data, we showed how the placement of FUSE in school allowed for *equitable access* and how challenges were designed (and redesigned) to appeal to both boys and girls. Using ethnographic data, we showed how FUSE promoted *equitable participation* in interest development and learning that was enactive, embodied, embedded, and extended. We showed how FUSE afforded opportunities for girls to organize their own learning, to become leaders of others’ learning, and to develop and become recognized for their expertise. We argue that these opportunities were afforded by three key features of the FUSE activity system. The first is the choice-based nature of FUSE, which allows learners to pursue projects of interest in ways that work for them. The second is the attention paid to designing and improving challenges so that they appeal to both girls and boys. The third is the way in which students are encouraged to help one another and seek help from peers, rather than relying on the teacher for support. These features suggest guidelines for the design of other equitable CSCL environments.

The analyses we’ve presented also provide guidelines for further investigations into other sorts of equity (or inequity) in the FUSE environment. For example, a similar analysis is needed to determine whether interest development and learning outcomes are equitably available to students of different racial, ethnic, and socioeconomic backgrounds. Our video-ethnographic observations suggest that they are. For example, three of the five girls represented here were Latinas, and one was Asian. However, a systematic examination of these forms of equity is needed. As FUSE is now scaling up to schools across the country, many of which serve majority low income and/or underrepresented minority populations, this has become an emerging focus of our research.

## References

- Atit, K., Ramey, K., Bar-El, D., Jona, K., Worsley, M., Stevens, R., Flanagan-Hall, G., Wu, J., Vondracek, M., & Uttal, D. (in press). Changing who is making: Broadening participation in maker activities. Proceedings of the First Annual Connected Learning Summit, Cambridge, MA.
- American Association of University Women (2000). *Tech-savvy: Educating girls in the new computer age*. Washington D.C.: American Association of University Women Educational Foundation. Commission on Technology, Gender, and Teacher Education.
- Barniskis, S. C. (2014). STEAM: Science and Art Meet in Rural Library Makerspaces. In *iConference 2014 Proceedings* (p. 834–837). doi:10.9776/14158
- Patton, M. Q. (1990). *Qualitative evaluation and research methods*. SAGE Publications, Inc.
- Bowler, L. (2014). Creativity through “Maker” experiences and design thinking in the education of librarians. *Knowledge Quest: Journal of the American Association of School Librarians*. 42(5), 59-61.
- Brahms, L. J. (2014). Making as a learning process: Identifying and supporting family learning in informal settings (Doctoral dissertation, University of Pittsburgh).

- Buchholz, B., Shively, K., Peppler, K., & Wohlwend, K. (2014). Hands on, hands off: Gendered access in crafting and electronics practices. *Mind, Culture, and Activity*, 21(4), 278-297
- Buechley, L., Eisenberg, M., Catchen, J., & Crockett, A. (2008, April). The LilyPad Arduino: using computational textiles to investigate engagement, aesthetics, and diversity in computer science education. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 423-432). ACM.
- DiGiacomo, D. K. & Gutiérrez, K. D. (2016). Relational equity as a design tool within making and tinkering activities. *Mind, Culture, and Activity*, 23(2), 141-153. doi:10.1080/10749039.2015.1058398
- Fields, D. A. & King, W. L. (2014). "So, I think I'm a programmer now." Developing connected learning for adults in a university craft technologies course. Proceedings of the 11<sup>th</sup> International Conference of the Learning Sciences, Boulder, CO: Volume 1, pp. 927-936.
- Goodwin, C. (2000). Action and embodiment within situated human interaction. *Journal of Pragmatics*, 32, 1489-1522.
- Hall, R., & Stevens, R. (2015). Developing approaches to interaction analysis of knowledge in use. In diSessa, A. A., Levin, M., & Brown, N. J. S. (Eds.), *Knowledge and interaction: A synthetic agenda for the learning sciences*. New York, NY: Routledge.
- Jordan, B. & Henderson, A. (1995). Interaction analysis: Foundations and practices. *The Journal of the Learning Sciences* 4(1), 39-103.
- Kafai, Y., Fields, D., & Searle, K. (2014). Electronic textiles as disruptive designs. *Harvard Ed Rev*, 84, 532-556.
- Kafai, Y., Searle, K., Martinez, C., & Brayboy, B. (2014). Ethnocomputing with electronic textiles. In *Proc of the 45th ACM SIGSCE* (pp. 241-246). ACM.
- Lachney, M., Babbitt, W. & English, R. (2016). Software Design in the 'Construction Genre' of Learning Technology: Content Aware Versus Content Agonistic. *Computational Culture*, 5.
- Lewis, J. (2015). Barriers to women's involvement in hackspaces and makerspaces. Accessed at <https://access-space.org/wp-content/uploads/2017/04/Barriers-to-womens-involvement-in-hackspaces-and-makerspaces.pdf>
- Margolis, J. & Fisher, A. (2002). *Unlocking the Clubhouse: Women in Computing*. Cambridge, Massachusetts: The MIT Press.
- Penney, L. R. (2016). An investigation into how students select and develop resources for learning as they pursue choice-based STEAM challenges (Doctoral dissertation, Northwestern University).
- Pinkard, N., Erete, S., Martin, C. K., & McKinney de Royston, M. (2017). Digital Youth Divas: Exploring narrative-driven curriculum to spark middle school girls' interest in computational activities, *Journal of the Learning Sciences*, 26(3), 477-516.
- Ramey, K. E. (2017). FUSE Studios: Bringing interest-driven, integrated STEAM learning into schools via makerspaces. (Doctoral dissertation, Northwestern University).
- Ramey, K. E. & Stevens, R. (submitted). Interest development and learning in choice-based, in-school, making activities: The case of a 3D printer. *Learning, Culture, and Social Interaction*.
- Ramey, K. E., Stevens, R., & Uttal, D. (submitted). In-FUSE-ing STEAM learning with spatial reasoning: Distributed spatial sensemaking in school making activities. *Journal of Educational Psychology*.
- Scott, K. A., Sheridan, K. M., & Clark, K. (2015). Culturally responsive computing: a theory revisited. *Learning, Media and Technology*, 40(4), 412-436.
- Stevens, R., Jona, K., Penney, L., Champion, D., Ramey, K. E., Hilppö, J., Echevarria, R., & Penuel, W. (2016). FUSE: An alternative infrastructure for empowering learners in schools. Proceedings of the 12<sup>th</sup> International Conference of the Learning Sciences, Singapore, SG, Volume 2, 1025-1032.
- Volman, M., & van Eck, E. (2001). Gender equity and information technology in education: The second decade. *Review of Educational Research*, 71(4), 613-634.
- Vossoughi, S. & Bevan, B. (2014). Making and tinkering: A review of the literature. *National Research Council Committee on Out of School Time STEM*, 1-55.
- Vossoughi, S., Escudé, M., Kong, F., & Hooper, P. (2013). *Tinkering, learning & equity in the after-school setting*. Paper presented at FabLearn, Stanford, CA. Retrieved from <http://fablearn.stanford.edu/2013/papers/>

## Acknowledgments

This work was supported by the National Science Foundation (NSF) grants DRL-1348800 and DRL-1433724. However, any opinions, findings, conclusions, or recommendations are our own and do not reflect the views of the NSF. We thank Jaakko Hilppö, Dionne Champion, Peter Meyerhoff, Lauren Penney, Kemi Jona, Henry Mann, Lucy Yang and the other members of the FUSE research and design team, as well as all our partner schools.

# Reading for Breadth, Reading for Depth: Understanding the Relationship Between Reading and Complex Thinking Using Epistemic Network Analysis

Hanall Sung, University of Wisconsin—Madison, hanall.sung@wisc.edu  
Shengyang Cao, University of Wisconsin—Madison, scaotravis@gmail.com  
Andrew R. Ruis, University of Wisconsin—Madison, arruis@wisc.edu

David Williamson Shaffer, University of Wisconsin—Madison and Aalborg University Copenhagen,  
dws@education.wisc.edu

**Abstract:** This paper examines whether and to what extent long and short readers make different contributions to collaborative design discussions in a CSCL environment—that is, we investigate whether a simple measure of reading behavior based on clickstream data is a good proxy for engagement with readings. Our approach to addressing this question is multimodal, involving two sources of data: (a) a record of students' online conversations, and (b) the frequency and duration with which documents were open on each student's screen. This study suggests that in this specific case, relatively thin data about reading frequency and mean reading duration can be used to make inferences about students' reading behavior in a CSCL context where it is impossible to observe directly. It also shows the power of a multimodal approach to the data—here, we are using one mode of data (discussion) to get a better understanding of another mode (clickstream).

## Introduction

In computer-supported collaborative learning (CSCL) environments, people learn complex thinking through interactions with people and tools (Hutchins, 1995). However, there are important actions and interactions that cannot be observed. One particularly difficult activity to investigate in CSCL contexts is reading. Prior research on text comprehension (Bell, 2001) suggests that there is a correlation between the amount of time that a student reads and the extent to which they understand the text. However, these studies are based on observations of the time students spend actually attending to a document. While it is possible to use eye-tracking systems to model what parts of a text a student focuses on in a CSCL context, this requires equipment that is difficult to deploy at scale (Rayner, Chace, Slattery, & Ashby, 2006). An alternative approach is to infer from clicking and scrolling behavior how a student engages with a text.

But how much data about student reading is needed to understand what role text resources play in a CSCL environment? The answer to that question clearly depends on the specifics of the CSCL environment: what is being learned, by whom, and through what activities, and what role the information from texts plays in the process. In this paper, we argue that it may be possible to infer students' level of engagement with a text using only a small amount of information: the frequency with which students open documents, and the length of time that the documents remain open on the screen. Most CSCL environments already record these as part of their *clickstream data*—the data collected by the CSCL system as students work. Using one CSCL environment, we show that this relatively sparse data is a reasonable proxy for students' depth of engagement with texts.

Specifically, we investigate student learning in the context of a *virtual internship* in which students read engineering resources and collaborate with other students through an online chat interface to design an assistive mechanical device. Most of the information that is the basis for collaborative discussions and design experimentation is conveyed by technical documents and research reports. In other words, reading plays a central role in this context. There is thus an important interaction between students' engagement with the readings and their collaborative discussions, both of which influence how they make design decisions.

Our approach to addressing this question is multimodal. We used clickstream data to identify (a) *long readers* (low frequency, high duration) and (b) *short readers* (high frequency, low duration). We then used chat data to examine whether and to what extent long and short readers made different contributions to collaborative discussions—that is, we investigated whether a simple measure of reading behavior based on clickstream data could be used to make inferences about students' depth of engagement with the texts in this CSCL setting.

To accomplish this, we used epistemic network analysis (ENA; Shaffer, Collier, and Ruis, 2016) to analyze students' discourse in the domain based on their contributions to collaborative discussions. We then used the resulting ENA model to investigate the differences in discourse patterns between long and short readers. Our results show meaningful differences in discourse between long and short readers, suggesting that a thick stream of data (such as chats) can be used to interpret the meaning of a thinner stream of data (such as

information on when documents are opened and closed). That is, in addition to using multiple data sources to get a better understanding of student learning, we used one mode of data (discussion data) to get a better understanding of another mode (clickstream data).

## Theory

A broad range of work in CSCL has shown that analyses of learning should not focus on a (hypothetical) unassisted individual, but rather need to consider individuals collaborating with others and using artifacts to solve complex problems (Hutchins, 1995; Lave, 1988; Shaffer, 2017). One particularly important tool for learning is written texts, which are a prominent feature of many learning environments, both computer-supported and face-to-face (Snow, 2002). However, investigating reading engagement in a CSCL context is difficult because we cannot observe students directly (Siemens, 2013). Some researchers attempt to use eye-tracking to model engagement with texts, but such studies are difficult at scale and require considerable expertise to analyze the data (Rayner et al., 2006).

Clickstream data, which most CSCL environments record, could also provide evidence of engagement with items on screen. For example, Coiro (2003) argues that *reading duration*—the amount of uninterrupted time on which a text appears on screen—can be used as a proxy measure for *engagement* with a text. However, the relationship between reading time and reading comprehension is complex. On one hand, researchers generally agree that more time spent reading leads to improvement in reading comprehension (McKeown, Beck, & Blake, 2009). On the other hand, the exact nature of the relationship is still debated (Bell, 2001). For example, Coiro (2003) argues that there are two basic reading behaviors: (a) *skimming*, or “getting only the gist of text in a short time,” and (b) *studying*, or “reading texts with the intent of retaining the information for a period of time.”

Studying behavior, however, is reflected not only in the *total duration* of time that a student engages with a reading, but also in whether they engage for sufficient time in one reading session to process the information. This is distinct from a student who might spend the same total amount time on reading but do so in a series of small and separate chunks, which is more indicative of skimming behavior. If we only consider reading duration as a metric of reading engagement, it would be difficult to distinguish between someone who opens a document once and spends a period time attempting to understand the contents, and someone who opens a reading multiple times, but spends only a short amount of time in each case. Similarly, it is difficult to distinguish students who spend a long time reading because they are engaging deeply with the content and those who spend a long time reading because they are struggling to understand it. Long reading duration, especially combined with repeated reading of the same text, can indicate students who are struggling to retain information in sufficiently large amounts to enable full comprehension (Bell, 2001; Rasinski, 2000). Thus, we need to take into account both the frequency and duration of reading to distinguish skimming behavior from studying behavior in a CSCL context.

Of course, we are ultimately interested in the extent to which reading behaviors contribute to understanding more generally. Research (Snow, 2002) shows that students who have low reading comprehension develop only very shallow knowledge, such as scattered facts and simple definitions of key terms. Shallow understanding is not sufficient to solve complex problems or apply knowledge to new situations. This suggests that different reading comprehension levels may relate to different levels of understanding, and thus different ways of framing, investigating, and solving complex problems. To model distributed cognition in the context of reading, then, we need to model *how students apply knowledge from what they have read to their collaborations with others* in order to solve complex problems.

To understand this complex relationship between reading and collaborative problem solving, we chose to analyze data from a group of students solving an engineering design problem in a CSCL simulation that positioned them as interns at an engineering firm. That is, we examined the interplay between reading and collaborative discussion in the context of an authentic learning task (Shaffer, 2006).

Shaffer (2012) argues that application of knowledge to a real-world problem in the context of a community of practice like engineering involves the development (and deployment) of an *epistemic frame*: a pattern of connections among knowledge, skills, and other cognitive elements that characterize a discourse community. Here we model how students have understood what they read by seeing whether and how they integrate the readings into the knowledge, skills, and other elements of practice in the domain (Collier, Ruis, & Shaffer, 2016).

These relationships can be modeled with ENA, a technique for identifying and quantifying connections among epistemic frame elements and representing them in dynamic network models. Critically, ENA accounts for both connections made within an individual’s own discourse and connections made to the discourse of other individuals in a collaborative discussion. That is, ENA models how an individual contributes to collaborative



discourse. In this study, we explore whether long readers have different epistemic frames than short readers by comparing their ENA networks, which indicate the contributions each student made to collaborative problem-solving activities. Specifically, we ask:

RQ1: Do long readers and short readers show different patterns of interaction in collaborative discussions?

RQ2: Do these differences (if any) reflect a difference in the depth to which long and short readers are engaging in the text?

## Methods

### Research context

We analyzed data from a virtual internship, *RescuShell*, in which students play the role of an intern at a fictional engineering design firm. In *RescuShell*, students design the robotic legs for a mechanical exoskeleton to be used by rescue workers. They use an online work portal with text resources, simulated design tools, and a built-in chat interface to collaborate with their project teams. The virtual internship simulates the engineering design process, including reviewing and summarizing research reports, creating device prototypes, discussing design choices with teammates, and working to balance the needs of various stakeholders (Chesler et al., 2015).

In *RescuShell*, the primary source of information about the technical constraints and performance parameters is a set of reading materials, including technical reports and research briefs. These documents, which consist of detailed text descriptions as well as tables, graphs, and images, help students gain sufficient technical knowledge to design and evaluate the performance of a mechanical exoskeleton. Importantly, key information is not concentrated in one place, but is diffused across the documents. That is, students need to integrate information from various documents and then discuss design decisions with their teammates, providing insight into how and to what extent the readings inform their design reasoning.

In this study, we analyzed group discussions (12,859 lines of chat data) and individual clickstreams (24,034 lines of clickstream data) from 203 college students who used *RescuShell* at eight different sites in the United States between 2013 and 2015.

### Data analysis

#### Identifying long and short readers

We extracted students' clickstream data, which consist of a time-stamped record of clicks in the system, including accessing resources, sending emails and messages, saving notes, and so on. We then calculated (a) how often each student opened one of the 17 different documents (frequency), and (b) the length of time each document was open on screen before the student clicked on anything outside the document (duration).

The documents in *RescuShell* contain an average of 463 words, with a range 153 to 786 words. Although reading speed varies based on a number of factors, college students typically read around 450 words per minute (Carver, 1992). In this study, the median frequency was 104 clicks (i.e., accessing the readings 104 times over the course of the virtual internship), and the median duration was 48 seconds. We used these values to divide students into long (lower frequency, higher duration than the medians) and short (higher frequency, lower duration than the medians) reading groups. Students with both frequencies and mean durations lower than the medians were omitted from the analysis, as this indicates little meaningful engagement with the readings. Students with frequencies and mean durations higher than the medians were also omitted, as this could indicate students who were struggling to understand the content.

#### Discourse coding

Chat transcripts were segmented by utterance, defined as when a student sent a single message in the chat program. To code the chat data for key epistemic frame elements relevant to the simulated engineering design process, we used an automated coding process (ncodeR; Marquart, Swiecki, Eagan, & Shaffer, 2018) based on regular expression matching. We validated all six codes using a series of comparisons between two human raters and ncodeR; pairwise Cohen's kappa scores ranged between 0.83 and 1.00 (see Table 1). We used Shaffer's rho to determine, for each kappa value, the likelihood that it would be found by two coders if their true rate of agreement was  $\kappa < 0.65$  (Shaffer, 2017). As shown in Table 2, all of the kappa values have rho values less than 0.05, meaning that if the coders were to code the whole dataset, they would have a level of agreement of  $\kappa > 0.65$  with a Type I error rate of less than 5%.

Table 1: Coding scheme and inter-rater reliability statistics (\*indicates  $p(0.65)<0.05$ ; \*\*indicates  $p(0.65)<0.01$ )

| Code Name                      | Description  | Example   | Kappa R1 v. R2 | Kappa R1 v. ncodeR | Kappa R2 v. ncodeR |
|--------------------------------|--|---|----------------|--------------------|--------------------|
| DESIGN REASONING               | Design development, prioritization, tradeoffs, and decisions                                     | <i>“Steel can carry a big load, but it is heavy and weighs down on the recharge interval, and it is a costly option.”</i> | 0.89**         | 0.89*              | 0.89**             |
| PERFORMANCE PARAMETERS         | Attributes: payload, recharge interval, agility, safety, or cost.                                | <i>“My device has a pretty good safety, payload, agility, and recharge interval”</i>                                      | 0.89**         | 1.00**             | 0.89**             |
| TECHNICAL CONSTRAINTS          | Inputs: actuators, ROM, materials, power sources, or sensors.                                    | <i>“Our two best were both made with Aluminum, NiCd Batteries, Piezoelectric sensors, and Pneumatic actuators.”</i>       | 0.83**         | 0.94**             | 0.89**             |
| CLIENT AND CONSULTANT REQUESTS | Decisions based on internal consultant’s requests or client’s health or comfort                  | <i>“We tried to meet at least the minimum of each of the internal consultant’s request.”</i>                              | 1.00**         | 1.00*              | 1.00**             |
| COLLABORATION                  | Facilitating a joint meeting or the production of team design products.                          | <i>“How should we make our team batch?”</i>   | 1.00**         | 1.00*              | 1.00**             |
| DATA                           | Discussion of numerical values, results tables, graphs, research papers, or relative quantities. | <i>“I thought that safety near the maximum was not very good (close to 225 - one had 218 RPN)”</i>                        | 0.90**         | 0.87**             | 0.89**             |

### ENA Discourse Model

To construct the ENA model, we defined the units of analysis as all lines of data associated with a single student. The ENA algorithm uses a moving window to construct a network model for each line in the data, showing how codes in the current line are connected to codes that occur within the recent temporal context (Siebert-Evenstone et al., 2017). Based on a grounded analysis of the data, we used a window of 7 lines (each line plus the 6 previous lines) within each team activity. The resulting networks are aggregated for all lines for each unit of analysis in the model. Networks were normalized to account for the fact that some students spoke more than others. We used a dimensional reduction that placed the means of the groups of long and short readers as close as possible to the x-axis of the projected space, and the y-axis was defined by the first dimension of a singular value decomposition (Shaffer, Collier, & Ruis, 2016).

ENA networks were visualized using network graphs where the nodes correspond to the codes, and the edges reflect the relative frequency of co-occurrence, or connection, between two codes. ENA produces two coordinated representations for each unit of analysis: (1) a plotted point, which represents the location of that unit’s network in the projected space, and (2) a weighted network graph. The positions of the network graph nodes are fixed, and determined by an optimization routine that minimizes the difference between the plotted points and their corresponding network centroids. Because of this co-registration of network graphs and projected space (co-registration correlations (Spearman) were dimension 1 = 0.93, dimension 2 = 0.96), the positions of the network graph nodes—and the connections they define—can be used to interpret the dimensions of the projected space and explain the positions of plotted points in the space. To test for differences between the networks of long and short readers, we applied a two-sample t-test, assuming unequal variance to the location of points in the projected ENA space, then used the corresponding network graphs to interpret the statistically significant differences.

## Results

### Quantitative results

Figure 1 shows each student’s network location, along with the means and 95% confidence intervals of the long readers (blue) and the short readers (red). There is a statistically significant difference between long and short readers on the first dimension with a moderate effect size ( $mean_{long} = 0.02$ ,  $mean_{short} = -0.18$ ;  $t = 3.28$ ,  $p < 0.01$ , Cohen’s  $d = 0.57$ ).

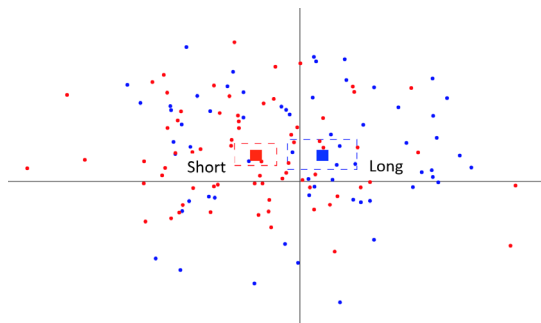


Figure 1. ENA scatter plot showing long (blue) and short (red) readers. Each point is a single student; the squares are group means; the dashed boxes are 95% confidence intervals (t-distribution).

To examine which connections accounted for the differences between long and short readers, we constructed mean epistemic networks for each group. As Figure 2 shows, both long readers (blue network, right) and short readers (red network, left) made dense networks of connections. This suggests that students in both groups were engaging in important engineering design practices. However, when we subtract one network from the other (Figure 2, middle) to identify why there is a statistically significant difference between the two groups, the long readers (blue network, right) made more links between DATA and other elements of the design reasoning process in their discussions, while short readers (left) made more links FROM TECHNICAL CONSTRAINTS to the other elements of the design reasoning process.

Note that the two groups were not having independent discussions: 58% of the discussions involved both long and short readers, and these results did not differ between mixed and homogenous groups.

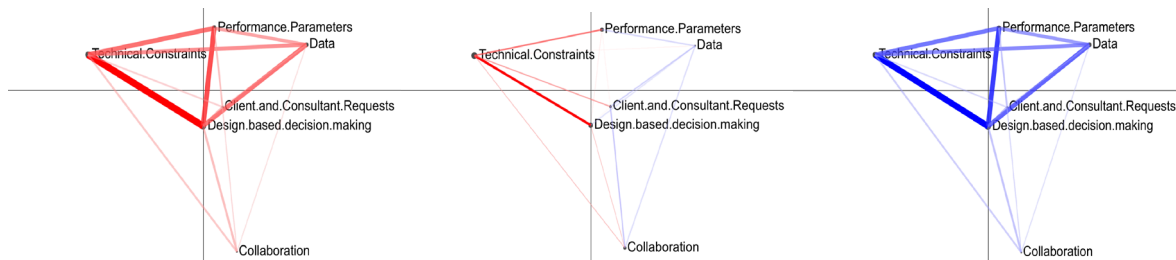


Figure 2. Mean ENA network diagrams showing the connections made by long readers and short readers. Short readers (left, red) mainly connected TECHNICAL CONSTRAINTS with other elements; long readers (right, blue) primarily connected DATA with design reasoning elements.

### Qualitative results

To understand the importance of these differences between long and short readers, we analyzed students' discussions qualitatively. Here, we present an example of a student discussion in *RescuShell*. This discussion took place at an early stage in the design process, when students meet to decide on a power source and control sensors for an experimental prototype. Just before the meeting, students read 6 technical documents, which are their only source of information about the different power sources and control sensors. For example, the *Safety Standard Handbook* provides data about the probability of failure modes occurring for different batteries and control sensors. The documents include detailed descriptions of how different sensors function, as well as numeric data about the effects of the sensors on payload, agility, battery recharge interval, cost, and safety. The *Control Sensor Overview* contains the table shown below (Table 2), which indicates how much each sensor costs and how each sensor affects device safety. (In this case, the higher the risk priority number, the less safe the device will be.) However, no single document contains complete information about the effects of different power sources and control sensors on exoskeleton performance. Thus, integrating information from multiple documents is the only way for students to understand the tradeoffs associated with different design choices.

Table 2: Data table provided in the *Control Sensor Overview*

| Control Sensor       | Strain-Gauge | Piezoelectric | Optic-Binary |
|----------------------|--------------|---------------|--------------|
| Cost per Sensor (\$) | 99           | 110           | 54           |
| Risk Priority Number | 84           | 55            | 102          |

After reading these documents, students meet with their project teams to discuss which control sensor they should choose for their design. Table 3 provides an excerpt from one project team’s conversation about which control sensor to choose. In the conversation are Sam and Mike, both identified as short readers based on their clickstream data, and Joe, identified as a long reader.

Table 3: Excerpt of one project team’s discussion about choosing a control sensor

| Line | Student | Reader Type | Discussion Utterance  | Codes   |
|------|---------|-------------|---|---|
| 1    | Sam     | Short       | So what attributes for the sensor does everyone think is most important? Then maybe we can choose a sensor based on that.   | TECHNICAL CONSTRAINTS   |
| 2    | Joe     | Long        | Sounds good, we can think about the design later. As for the sensor I would still go with the Piezoelectric sensor because even though it is the most expensive, it is such only by \$11, it has the highest recharge interval, second highest agility. | DESIGN REASONING<br>PERFORMANCE PARAMETERS<br>TECHNICAL CONSTRAINTS<br>DATA |
| 3    | Mike    | Short       | Looking at the data table, I would choose the optic binary sensor. However, this does not take into account the need for reflection in order for it to work. I agree with Joe and choose the Piezoelectric.   | DESIGN REASONING<br>TECHNICAL CONSTRAINTS<br>DATA                           |
| 4    | Joe     | Long        | I don't think we should use the optic binary sensor because although the cost is pretty cheap, and it has the best recharge interval by about 40 minutes, it performs worst in agility and safety.  | DESIGN REASONING<br>PERFORMANCE PARAMETERS<br>TECHNICAL CONSTRAINTS<br>DATA |

**Line 1: Sam’s Question:** In line 1, Sam asks his teammates: “what attributes for the sensor does everyone think is most important?” Thus, he suggests that identifying PERFORMANCE PARAMETERS will help the team decide which sensor to choose.

**Line 2: Joe’s Proposal:** Joe, a long reader, begins line 2 by declaring his preference: “I would still go with the Piezoelectric sensor” (TECHNICAL CONSTRAINTS). However, he does not just state his preference, he goes on to *justify* his claim (that is, use DESIGN REASONING). First, he acknowledges, the sensor he chose has a downside (“it is the most expensive”; PERFORMANCE PARAMETERS) but he also addresses this concern by arguing that it is only the most expensive “by \$11” (DATA). Finally, he explains that his choice has two highly desirable attributes: “the highest recharge interval, second highest agility.” Thus, he is proposing and justifying his design choice by weighing the pros and cons of his proposal (DESIGN REASONING).

In doing this, Joe is integrating information from multiple sources. The comparative cost information among control sensors comes from the data table in the *Control Sensor Overview* (shown above). This document explains the relationship between piezoelectric sensors and two important PERFORMANCE PARAMETERS: recharge interval and agility. However, the *ranking* of the recharge interval and agility among three control sensors—which is the basis of Joe’s justification—can only be found in graphs from a different document: the *Control Sensor and Power Source Experimental Report*, which describes the results obtained from experiments to evaluate the recharge interval and agility of three different control sensors. In order to make justify his choice in terms of both performance parameters and cost, in other words, Joe used pieces of information from different documents and synthesized them into a coherent argument.

**Line 3: Mike’s response:** In line 3, Mike, one of the short readers, also references the data table from the *Control Sensor Overview* (“looking at the data table...”; DATA) shown in Table 2. And he also states his preference: “I would choose the optic binary sensor” (TECHNICAL CONSTRAINTS). However, unlike Joe, Mike does not provide a clear justification for choosing the optic binary sensor. Instead, he provides an *explanation* of how the optic binary sensor functions in contrast to the piezoelectric: “This does not take into account the need for reflection in order for it to work” (DESIGN REASONING). If we refer back to Table 2 from the *Control Sensor Overview*, we might *infer* that he was thinking about the cost efficiency: the optic binary sensor has the lowest cost. But he does not actually talk explicitly about any of the performance attributes—which were the basis of Sam’s original question. Instead, he refers to an explanation of how optic binary sensors function that also comes from the *Control Sensor Overview*. Based on this limitation, Mike accepts Joe’s proposal.

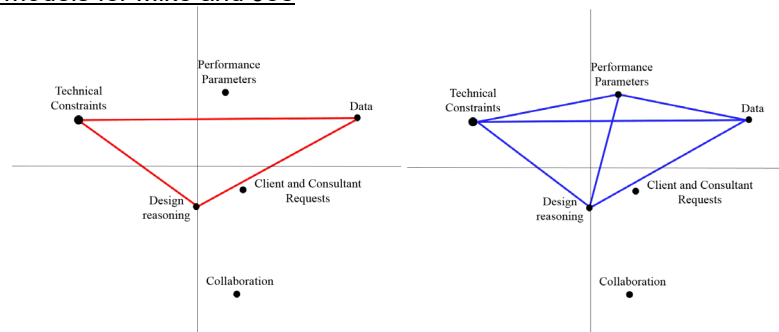
Thus, Mike is clearly engaged with the *Control Sensor Overview* document: he references the data table and knows how the optic binary sensor works. However, all of this information comes from the same

document. He is not integrating information across multiple readings, and he is not reading deeply enough to recognize that in this situation, his explanation of the mechanisms by which the binary optical sensor functions is less relevant than its impact on the PERFORMANCE PARAMETERS of the device.

**Line 4: Joe’s response:** In line 4, Joe disagrees with Mike: “I don’t think we should use the optic binary sensor” (TECHNICAL CONSTRAINTS), and he explicitly refers to four PERFORMANCE PARAMETERS, where Mike did not refer to any. He recognizes the optic binary sensor has benefits on cost (“the cost is pretty cheap”) and recharge interval (“it has the best recharge interval by about 40 minutes”; DATA). However, he emphasizes that the optic binary “performs worst in agility and safety” (DESIGN REASONING).

That is, Joe reuses the cost and safety information from the *Control Sensor Overview* shown in Table 2 and combines the information from the *Control Sensor and Power Source Experimental Report* on recharge interval, payload, and agility, and the *Safety Standard Handbook* on safety information, to determine the impact of optic binary sensors. Once again, he is looking across multiple documents and *integrating* information about multiple performance attributes to support his argument against an alternative to his original proposal.

### Comparing ENA models for Mike and Joe



**Figure 3.** Individual ENA networks showing the connections made by Joe, the long reader (right) and Mike, the short reader (left). Joe connected DATA with TECHNICAL CONSTRAINTS, DESIGN REASONING, and PERFORMANCE PARAMETERS; Mike (left) connected only TECHNICAL CONSTRAINTS with DATA and DESIGN REASONING.

Figure 3 above shows the individual networks of Mike and Joe from just this short excerpt. In line 2, when Joe talks about his preference, he makes a connection between TECHNICAL CONSTRAINTS and DESIGN REASONING when he justifies his choice. He also makes a connection between PERFORMANCE PARAMETERS and DATA by explaining his choice in terms of both performance parameters and cost with actual value from multiple documents. Similarly, in line 4, Joe makes connections among TECHNICAL CONSTRAINTS, DATA, PERFORMANCE PARAMETERS, and DESIGN REASONING. On the other hand, Mike makes a connection between TECHNICAL CONSTRAINTS and DESIGN REASONING (line 3), and TECHNICAL CONSTRAINTS and DATA (line 3). These individual networks of Mike and Joe align with the characteristics of long and short readers’ group networks in the sense that they show Joe linking concepts more completely than Mike did, and specifically making more robust connections to DATA.

Combined with the qualitative analysis, which shows that linking of these concepts is a reflection of deeper engagement with the readings, and the quantitative analysis above, which shows that longer readers consistently make more robust use of data during design discussions, suggests that longer reading is associated with deeper reading.

## Discussion

In this paper, we investigated whether long readers make different contributions to collaborative discussions than short readers in one CSCL context. Our results show that short readers were less likely to be able to articulate complex arguments with clear justifications, whereas long readers who engage more deeply with the readings were better able to flexibly and dynamically integrate information from multiple sources and work it into their arguments in collaborative discussions. Given these differences, our findings suggest that in this case, relatively thin data about reading frequency and mean reading duration could be used to make inferences about students’ reading behavior in a CSCL context where it is impossible to directly observe students’ reading behavior directly. It also shows the power of a multimodal approach to the data—and in particular, it shows that in addition to using multimodal data to get a better understanding of the student learning, we can also use one mode of data (in this case, discussion data) to get a better understanding of another mode (in this case, clickstream data).

This study has several limitations. First, it does not show directly that reading frequency and duration correspond with reading comprehension, only that they correspond with more or less sophisticated contributions to collaborative discussions. In particular, it did not model the relationship between specific reading behaviors and contributions to collaborative discussions in temporal context. Moreover, further research would be needed to disambiguate the effects of reading frequency and duration from other variables, such as prior engineering knowledge. In future work, we plan to build on this work to address these shortcomings by modeling the relationships between reading behaviors and discussion contributions as they occur in temporal proximity. Despite these limitations, this study suggests that student reading behaviors are associated with complex problem-solving behaviors: specifically, that long readers read more deeply, and are thus able to make more sophisticated contributions to collaborative problem-solving efforts than short readers who are reading shallowly. Moreover, it provides evidence that the frequency and duration of reading, which can be easily determined from the clickstream data recorded by most CSCL environments, can in some cases be used as a proxy for reading engagement, which is difficult to observe directly in virtual settings.

## References

- Bell, T. (2001). Extensive reading: Speed and comprehension. *The reading matrix*, 1(1).
- Carver, R. P. (1992). Reading rate: Theory, research, and practical implications. *Journal of Reading*, 36(2), 84-95.
- Chesler, N.C., Ruis, A.R., Collier, W., Swiecki, Z., Arastoopour, G., & Shaffer, D.W. (2015). A novel paradigm for engineering education: Virtual internships with individualized mentoring and assessment of engineering thinking. *Journal of Biomechanical Engineering*, 137(2).
- Coiro, J. (2003). Exploring literacy on the internet: Reading comprehension on the internet: Expanding our understanding of reading comprehension to encompass new literacies. *The reading teacher*, 56(5), 458-464.
- Collier, W., Ruis, A., & Shaffer, D. W. (2016). Local versus global connection making in discourse. *Paper presented at the 12th International Conference of the Learning Sciences*. Singapore.
- Hutchins, E. (1995). *Cognition in the Wild*. MIT press.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174.
- Lave, J. (1988). *Cognition in Practice*. Cambridge: Cambridge University Press.
- Marquart, C. L., Swiecki, Z., Eagan, B., & Shaffer, D. W. (2018). ncodeR: Techniques for Automated Classifiers. R package version 0.1.2. <https://CRAN.R-project.org/package=ncodeR>
- McKeown, M. G., Beck, I. L., & Blake, R. G. (2009). Rethinking reading comprehension instruction: A comparison of instruction for strategies and content approaches. *Reading Research Quarterly*, 44(3), 218-253.
- Rasinski, T. V. (2000). Commentary: Speed does matter in reading. *The Reading Teacher*, 54(2), 146-151.
- Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye movements as reflections of comprehension processes in reading. *Scientific studies of reading*, 10(3), 241-255.
- Shaffer, D. W. (2006). *How computer games help children learn*. Macmillan.
- Shaffer, D.W. (2012). Models of situated action: Computer games and the problem of transfer. In C. Steinkuehler, K. Squire, S. Barab (Eds.), *Games learning, and society: Learning and meaning in the digital age*, (pp. 403-433). Cambridge, UK: Cambridge University Press.
- Shaffer, D.W. (2017). *Quantitative Ethnography*. Madison, WI: Cathcart Press.
- Shaffer, D.W., Collier, W., & Ruis, A.R. (2016). A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics*, 3(3), 9-45.
- Siebert-Evenstone, A. L., Irgens, G. A., Collier, W., Swiecki, Z., Ruis, A. R., & Shaffer, D. W. (2017). In Search of Conversational Grain Size: Modelling Semantic Structure Using Moving Stanza Windows. *Journal of Learning Analytics*, 4(3), 123-139.
- Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist*, 57(10), 1380-1400.
- Snow, C. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Rand Corp.

## Acknowledgments

This work was funded in part by the National Science Foundation (DRL-1661036, DRL-1713110), the U.S. Army Research Laboratory (W911NF-18-2-0039), the Wisconsin Alumni Research Foundation, and the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison. The opinions, findings, and conclusions do not reflect the views of the funding agencies, cooperating institutions, or other individuals.

# On a Making-&-Tinkering Approach to Learning Mathematics in Formal Education: Knowledge Gains, Attitudes, and 21<sup>st</sup>-Century Skills

Stella Timotheou, Cyprus Interaction Lab, Department of Multimedia and Graphic Arts, Cyprus University of Technology, stella@cyprusinteractionlab.com

Andri Ioannou, Cyprus Interaction Lab, Department of Multimedia and Graphic Arts, Cyprus University of Technology, Research Center on Interactive Media, Smart Systems and Emerging Technologies (RISE), andri@cyprusinteractionlab.com

**Abstract:** Making and tinkering, as learning practices, have gained a lot of attention during the last decade, especially in STEAM education. Despite the significant interest in making-&-tinkering activities, most of the research has focused on implementations in non-formal and informal educational settings, with learners of older age (i.e., students in secondary school). The present study sought to investigate young students' knowledge gains, attitudes towards a making-&-tinkering approach to learning mathematics in formal education and the development of 21<sup>st</sup>-century skills as they engaged in a collaborative making-&-tinkering project using a variety of arts, crafts, and technological tools such as a physical robot. The making-&-tinkering approach involved making, tinkering, programming, and play in a group project integrated into the formal mathematics curriculum. Findings from the study suggest that young students can greatly benefit from such an approach. The study has demonstrated the applicability and value of such methods for young learners in formal educational contexts, with implications for future research and practice in the field.

## Introduction and theoretical background

There is a lot of recent discourse on learning in STEAM (Science, Technology, Engineering, Arts, and Mathematics) and how it is promoted through making, tinkering, programming, and play activities in various contexts and settings, including CSCL ones. STEAM education has been given a lot of attention during the last decade, and it has become an emphasis in many curriculums around the world (Vossoughi & Bevan, 2014). Increased pressure has been placed on teachers to improve performance in the above disciplines (Vossoughi & Bevan, 2014). This improvement can be achieved through new ways of teaching. Students who have been taught through traditional methods may have limited critical thinking ability as these methods emphasize only on the right answer to the problem through a determined process (Rode et al., 2015). In contrast to traditional methods, newer ones that support making, tinkering and play, set the significance of the process (and not only the end-result) and provide better ways to attract students with diverse thinking styles (Rode et al., 2015). In the present investigation, we adopt a making-&-tinkering approach to learning Mathematics (M), together with Science (S), Technology (T), Arts (A) and Engineering (E), in formal education with young students in elementary school, aged 8 to 9 years old.

Making is a process of creating something (Hsu, Baldwin, & Ching, 2017), and specifically, it is “the act of creating tangible artifacts” (Rode et al., 2015, p. 8). The movement of making, called “the maker movement”, as an active process of building, designing, and innovating with tools and materials to produce shareable artifacts, has gained enormous momentum during the last years (Papavlasopoulou, Giannakos & Jaccheri, 2017). Making is a learner-driven educative practice that supports learning, participation, and understanding (Vossoughi & Bevan, 2014). Tinkering as a part of making (Vossoughi & Bevan, 2014), is a problem-solving technique and learning strategy, which promotes a practice of improvement, and it is associated with experimentation and “trial and error” methods (Krieger, Allen, & Rawn, 2015). As Martinez and Stager (2013) indicate, making and tinkering involve a playful approach to solving problems through direct experience, experimentation, and discovery. Programming and physical computing are very often making and tinkering activities (Hsu et al., 2017), as students can build and rebuild their robot, make the program design, code and debug. Both in programming and in making and tinkering activities, the play is diffused. Play in this work is defined as a dynamic, active and constructive behaviour, and it's contiguous with the gamification of learning (Deterding, 2011; Ioannou, 2018). The connection between all the above concepts and learning in CSCL, including students' attitudes, are examined in this paper.

Both making and tinkering in education are not new ideas. They have their theoretical roots in Papert's constructionism, which is built upon Piaget's constructivism (Papavlasopoulou et al., 2017). Making is also noted by educators such as Froebel, Montessori and Dewey when they advanced practical, physical and playful learning

(Vossoughi & Bevan, 2014). As Vossoughi and Bevan (2014) indicated, the theories of Vygotsky, Lave and Wenger are related to making, too, in the notion that making can support learning and development.

The benefits of making in the learning process have been identified for many decades, thus during the last years a growing interest has been shown on making or “maker culture” as a philosophy or phenomenon (Papavlasopoulou et al., 2017), which should be examined more thoroughly through more empirical studies (Hsu et al., 2017). Making can change learners’ role from passive recipients to active ones, taking control over their own knowledge (Papavlasopoulou et al., 2017), which it is a vital characteristic of learner-driven pedagogies that are compounded in curriculums. Learning by doing gives learners the opportunity to engage in problem-solving, self-directed and collaborative work (Hsu et al., 2017) through the development of interests, identity and knowledge, as it may involve the use of physical materials in traditional crafts or hobby techniques (such as woodworking, cooking, etc.) or the use of technology. During making, learners are engaged in tinkering, which also builds on inquiry-based pedagogy and exploits learner-centered learning (Papavlasopoulou et al., 2017). According to Sheridan et al. (2014), tinkering supports learning and promotes 21<sup>st</sup>-century skills. As programming involves tinkering, educational robotics as learning tool can also benefits teaching and learning in STEAM education. Making, tinkering and programming are closely related to playful and gameful learning, as experimental and hands-on activities are involved. Robots can be used as a game in students’ hands and alongside with gameful design (Deterding, 2011; Ioannou, 2018), the learning outcomes can be beneficial.

Although a few researchers focus on the value of making, tinkering, programming, and play in learning (Krieger et al., 2015; Hsu et al., 2017; Martinez & Stager, 2013), studies which document the child’s learning experience in making-based activities are currently lacking. A growing number of efforts in making-enhanced activities is observed in non-formal or informal learning environments, such as workshops, libraries, museums, summer-school and after-school programs rather than in formal education. Yet, most of these efforts are practice-oriented in contrast to only few research-oriented studies (Chu, Angello, Saenz, & Quek, 2017). Moreover, typically such activities are done with students of older age e.g., secondary education, 12-18 years old (Papavlasopoulou et al., 2017).

In the present investigation, we adopted a making-&tinkering approach to learning Mathematics (M), together with Arts (A) and Engineering (E), in formal education with young learners. The study sought to investigate young learners’ knowledge, attitudes towards a tinkering-&-making approach to learning mathematic and their development of 21<sup>st</sup>-century skills, as they engaged in a group project using a variety of arts, crafts, and technological tools. The study sought to answer the following research questions:

*RQ1: What kinds of learning gains do young students experience during making-&tinkering activities?*

*RQ2: What are their attitudes towards this approach of learning in formal (math) education?*

*RQ3: How this approach seems to enact the development of 21<sup>st</sup>-century skills?*

## Methodology

The study adopted a mixed methods research design, relying on both quantitative and qualitative data collection and analysis to answer the research questions under investigation.

## Participants

The sample was composed of 18, 3<sup>rd</sup>-grade students (aged 8 to 9 years old), 12 girls and 6 boys. The students came from a small public primary school in a rural area in northeastern Mediterranean. The majority of the students came from families of middle socioeconomic status and education. The school was selected by convenience sampling, as the researcher was teaching at the school.

## Procedures

The research work took place in the participants’ regular classroom. The desks were set in a way so as students could work in six groups of three. The group formation constituted children from different cognitive, emotional and social behavior levels. Two groups used the classroom’s computers and four groups used tablets. On each group’s table, there were pencils, rubbers, rulers, variety of reused materials and an Edison robot. The 5-sessions intervention lasted four weeks as summarized in Table 1 (see also pictures in Figures 1, 2).

Table 1: The 5-sessions intervention

| Sessions | Activity | Description  |
|----------|----------|--|
| 1st      | craft    | -the problem is given<br>-brainstorming about how they could make their craft and materials they could use |



|                 |  |   |
|-----------------|--|---|
| 2 <sup>nd</sup> | math goals                                 | -students in groups create their craft<br>-small exhibition in the classroom of their “monster robot vacuum cleaners”<br>-students write a four-digit price for their monster<br>-they are taught about rounding to the nearest hundred<br>-they choose six of their monsters’ prices (the number six was given to limit the task) and try to round them to the nearest hundred |
| 3 <sup>rd</sup> | finish the artefacts and program the robot | -students work in three-members’ group<br>-they choose which monster is suitable to be set on the robot<br>-they find a way to put the monster on the robot<br>-they program their artefacts to move continuously   |
| 4 <sup>th</sup> | improve the programming                    | -groups look for ways to improve their programming, by making their robot moving continuously without getting out of the borders  |
| 5 <sup>th</sup> | prepare posters                            | - groups prepare their poster and organize their presentation   |

## Data collection and analysis

Data collection included knowledge pre- and post- tests, questionnaire on students’ attitudes, video data and focus groups data.

### Knowledge test

A knowledge test was used pre and post intervention to measure students’ acquisition of mathematical knowledge. The test included two exercises on the content of the math lesson. The same test was given a month later (i.e., as delayed-test) to measure the retention of knowledge. The tests were examined for face validity by an experienced teacher-colleague at the school. Another teacher-colleague examined and ensured that the lesson was relevant and in-line with the National Curriculum.

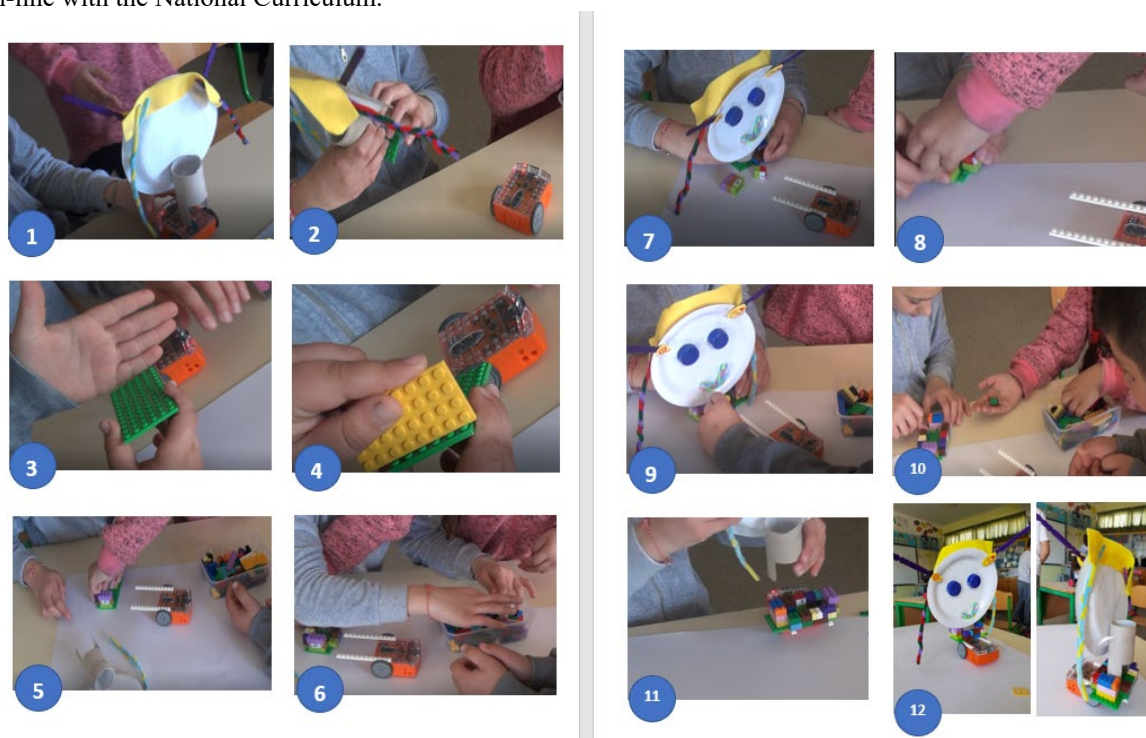


Figure 1. Making-&-tinkering activities to set the craft onto the robot.

### Attitudes questionnaire

The questionnaire used is known as the My Class Activities (MCA), an instrument developed by Gentry and Gable (2001) to measure students’ attitudes toward their class activities, specifically their level of: (a) interest e.g., “What I do in my class fits my interests”, “I have an opportunity to work on things in my class that interest me”, (b) challenge e.g., “The activities I do in my class are challenging”, “I have to think to solve problems in my class”, (c) choice e.g., “I can choose to work in a group”, “I can choose materials to work with in class”, and (d) enjoyment e.g., “I look forward to my class”, “I have fun in my class”. The questionnaire has previously reported good psychometric properties (Gentry & Gable, 2001). The questionnaire was translated from its English version

to Greek, following the standard test adaptation guidelines (Chapman & Carter, 1979). Initially the MCA items were translated by the researcher, who is a native speaker of Greek and then the items were translated back to English by an English teacher, and finally they were compared. Two more experts reviewed the translated MCA items, to ensure the content validity of the items. Then it was piloted with four students of the same school age, coming from another school of the same district. The MCA questionnaire was administered at the end of the 5<sup>th</sup> session of the intervention.

### Video data and analysis

Observational data were collected via video analysis, which was held in two rounds. In the first round, which was based on the Critical Incident Technique (Flanagan, 1954), general but important episodes and themes were noted (e.g. “Making and tinkering was detected during the phase of the craft’s construction”, “Tinkering helped students to improve their craft and their programming, to remember the result, and to feel satisfaction about their achievement”, “Collaboration during the making and tinkering activities helped them to improve their artifact and to find the best programming solution”). These themes helped in the organization of the focus groups’ protocol. The second round of video-analysis was conducted using the Manifest Content Approach (FitzGerald, 2012) on the episodes identified during the first round; the episodes were coded based on the Tinkering Dimensions Framework (Bevan et al., 2015) which was used as per Table 2. Both, the first and second video-analysis rounds were conducted separately by the researcher and another expert and inter-rater reliability (present agreement) was assessed to 87% and 89% respectively. Upon percent agreement computation, the raters discussed and resolved all disagreements.

Table 2: Learning Dimensions Framework developed by the Tinkering Studio (Bevan et al., 2015)

| <b>Learning Dimension</b>    | <b>Indicators</b>   |
|------------------------------|---|
| Engagement                   | <ul style="list-style-type: none"> <li>• Spending time in tinkering activities</li> <li>• Displaying motivation or investment through effect or behavior</li> </ul>   |
| Initiative & Intentionality  | <ul style="list-style-type: none"> <li>• Setting one’s own goals</li> <li>• Seeking and responding to feedback</li> <li>• Persisting to achieve goals in the problem space</li> <li>• <u>Taking intellectual risks or showing intellectual courage</u></li> </ul> |
| Social Scaffolding           | <ul style="list-style-type: none"> <li>• Requesting or offering help in solving problems</li> <li>• Inspiring new ideas or approaches</li> <li>• <u>Physically connecting to others’ works</u></li> </ul>   |
| Development of understanding | <ul style="list-style-type: none"> <li>• Expressing a realization through effect or utterances</li> <li>• Offering explanations for a strategy, tool or outcome</li> <li>• Applying knowledge</li> <li>• Striving to understand</li> </ul>                        |

### Focus groups

Semi-structured focus-group interviews with students (3 focus groups with 6 students each, 40 minutes each) were conducted at the end of the experience. The focus-group interviews were organized based on the framing of the study, the themes derived from the video analysis, and the research questions of the study (e.g., “What did you learn from the experience?”, “How do you feel about this approach to learning mathematic at school?”). The focus-group interview data were video-recorded and transcribed for subsequent data analysis. Data were transcribed and qualitatively analyzed using a thematic analysis approach (Attride-Stirling, 2001).

## **Findings**

### **Knowledge gains**

A Wilcoxon Signed Rank Test demonstrated statistically significant gains in students’ knowledge from pre- to post- testing ( $W(18) = 2, z = -2.18, p < .05$ ). Also, a Wilcoxon Signed Rank Test was used to examine knowledge retention (i.e., delayed test data); results showed that knowledge gains were preserved since there was no statistically significant difference from post to delayed testing ( $W(18) = 1, z = -1.34, p > .05$ ).

Consistent with the quantitative data, during the focus group discussions, almost all the children (n=15 out of 18) mentioned that they could explain how to round a four-digit number to the nearest hundred. Specifically, a girl noted:

*I really liked the session where we gave four-digit prices to our monster-hoovers and organized a small shop. And it was nice when we walked around the classroom -to the shop-desks- to choose 6 prices and try to round them to the nearest hundred...We didn't know how to round the numbers and we were trying [...] Only John was right and when he explained to us, I understood and now I know! [#12, Girl]*

### Attitudes towards the tinkering-&-making approach

Descriptive statistics on the attitude subscale mean-scores showed that the students appreciated and enjoyed the tinkering-&-making approach to formal (math) learning, particularly the opportunity to choose the way they could work in the project and to face challenges during the project, which triggered their interest (Table 3).

Table 3: Subscale mean-scores for attitudes

| Subscale  | Mean | SD   |
|-----------|------|------|
| Interest  | 3.30 | 0.61 |
| Challenge | 2.88 | 0.66 |
| Choice    | 1.90 | 0.77 |
| Enjoyment | 3.50 | 0.61 |

Findings from the video-analysis supported the results from questionnaire data on attitudes, as 27% of the incidents concerned the presence of the dimension of “Engagement” (see Figure 2), based in the framework by (Bevan et al., 2015).

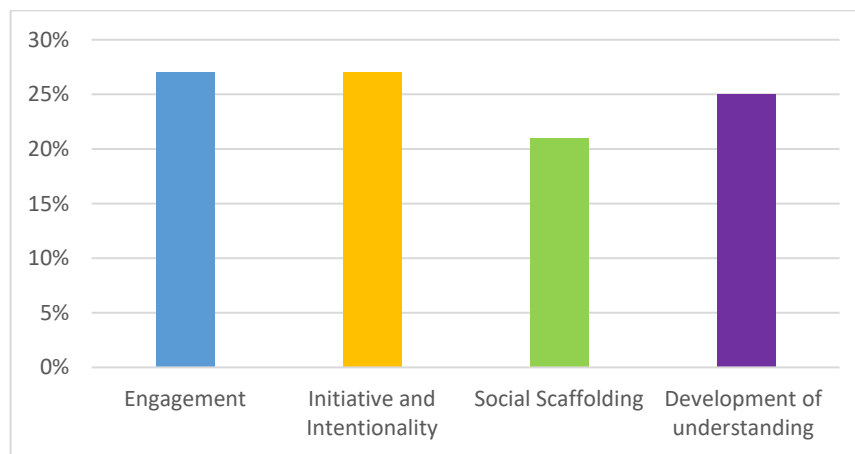


Figure 2. Results according to the "Learning Dimension Framework".

The focus-group interviews provided additional evidence of students' attitudes, in fact, demonstrating a positive shift in attitudes over time, e.g.

*At the beginning, we couldn't work [...] by the end, I liked it! When you gave us the programming challenge, we thought it was something difficult, but after we tried it was easier[...] I liked the project because we programmed the robots through the computers! [#1, Girl]*

*I was enthusiastic! Basically, we felt happiness! [...] All the sadness we felt at the beginning had gone away by the end as we were happy about our achievements! [...] That time I felt I was a scientist who invented something very great. And we learned a thousand of things we didn't know before! And using all those materials you gave us, we managed to do something we never did before[...] Will we make again a lesson with robots? [#4, Girl]*

## 21<sup>st</sup>-century skills

Most of the results derived from the focus-group interview data referred to the 21<sup>st</sup>-century skills. The data were grouped into three categories -- Learning & Innovation Skills, Information Media & Technology Skills, and Life & Career Skills – per framework by the Partnership for the 21<sup>st</sup> Century Learning (P21, 2015). Findings concerning the “Learning & Innovation Skills” are organized below in categories.

### Creativity and innovation

The focus-group data revealed that creativity and innovation skills were worked throughout the project. The students used their imagination to construct their craft and design and to prepare their poster. During the project, the students improved their crafts and their programming, invented new ideas, improved or changed their initial ideas and found ways to fasten their craft onto the robot, e.g.

*At the beginning, I had an idea, later I got another idea, and finally, I got another idea...  
We put our imagination to make our monsters and posters!  
I mostly liked the craft part and when I used my imagination!  
[...] and we managed to fasten our monster on the robot without any lego or anything else! We just set it on the robot and it wouldn't fall.  
My team found another way [...] we set the monster onto the robot, but when we realized that there was no space to put our fingers to press the robot's buttons, we used the small wood chips that were set on the monster to press the buttons. So, we turned the monster in a way that the yellow wood chip could press the circular button, the green wood chip could press the square button, and the other wood chip could press the triangle!!! [#13, Girl]*

In most groups, the students worked creatively with others and they implemented or suggested innovations, e.g.

*I liked the moment we invented new things.  
That time I felt I was a scientist who invented something very important!  
When you are engaged in such a kind of activities you feel like you are Einstein!  
Instead of the monster-hoover robot, we could have made the caveman robot related to our history [...]  
Yes! Or to use the idea in Gymnastic where the robots could be our bodies!"  
And invent more outdoor activities, not only indoor activities.  
[...] the teacher could use more puzzle challenges leading us to a treasure, for more mystery and fun!  
Oh! We could also write the programming commands on a paper and later when we find the treasure, we could test them on the tablet. [#18, Girl]*

### Critical thinking and problem-solving

In students' reporting we could easily identify strategies related to critical thinking and problem solving e.g.

*We learned how to describe step by step the process we followed from the beginning to the end  
[...] we planned with all the steps until we could solve the problem and reached our goal!  
[...] we made questions to find other solutions [...]  
We should be careful, and we should observe what we were doing to avoid making mistakes...  
We knew why we made each step, and we were able to explain it...  
At the beginning, we couldn't find the right programming solution, but later we understood how to do it! I know how to do it!" (talking about rounding to the nearest hundred)  
I realized that we must finish something before we review and make judgments [...] it may not be so difficult [...]  
why the other group found the programming and we did not?  
[...] I saw the other posters full of colours and I felt nervous. [#12, Boy]*

### Communication and collaboration

It became apparent in the focus-group interview data that the students communicated clearly with others via a variety of communicative ways, ranging from body movements and discussions to the usage of descriptions, posters and narrative stories. They also learned to collaborate with others especially their groupmates, e.g.

*At the beginning, when one child told us to do it like this, the other one disagreed and we started to quarrel each other, but later, we managed to collaborate [...]  
I would like to have more group works and collaborate with more kids.  
We learned to work in groups without loud voices and quarrels.  
We learned to respect each other in our group, especially we learned to help, and give advice [...]  
When my groupmate resigned and didn't want to try again in finding the right programming, I was telling her to continue and that we could make it! [#8, Boy]*

Even though most data were categorized as “Learning & Innovation Skills”, students also reported ideas related to “Information Media & Technology Skills”, e.g.

*We learned how to use the technology [...] our robot, the computer, the tablet, and we first saw and touched an iPod and a camera! [#3, Boy]*

*We managed to program our robot...We learned to use the computer or the tablet, and to program the robot...I learned robotics and programming! [#6, Boy]*

Furthermore, students also reported learning gains related to “Life & Career Skills”, e.g.

*The project taught me that I must have patience and persistence, because I had to finish something. [#2, Girl]*

*We were so happy, because we tried again and again, to make it [...] to what we wanted to... we never stopped trying. [#12, Girl]*

*We've been taught to respect others [...] and help each other [...] We've been taught to be hard-workers [...] to reach our goals! [#1, Girl]*

## Discussion and implications

The present study sought to investigate students' knowledge gains, attitudes towards a making-&-tinkering approach to learning mathematics and development of 21<sup>st</sup>-century skills as they engaged in a making-&-tinkering project using a variety of arts, crafts, and technological tools such as a physical robot. The approach involved making, tinkering, programming, and play in a multidisciplinary project integrated into the mathematics curriculum. Findings from the study suggest that young students can greatly benefit from such an approach. While making-&-tinkering activities are typically enacted in informal learning settings with learners of older age e.g., secondary education, 12-18 years old (Papavlasopoulou et al., 2017), the present study demonstrated the applicability and value of such methods, in formal educational context with young students.

*RQ1: What kinds of learning gains do young students experience during making-&-tinkering activities?*

Findings demonstrated that students' mathematical **knowledge** was improved by the end of the project and that the knowledge gained was retained, consistent with previous work reporting that making is in favour of knowledge acquisition (Chu et al., 2017). Indeed, previous work has also reported that making activities helped young children develop knowledge whilst the nature of these activities helped them share the new knowledge with others (Sheridan et al., 2014).

*RQ2: What are their attitudes towards the tinkering-&-making approach in formal (math) education?*

Findings from the present investigation suggest positive shifts in students' **attitudes** towards the making-&-tinkering approach. Interestingly, while at the beginning students exhibited negative behaviours, the implementation of the interdisciplinary project allowed shifts towards more positive attitudes; the shifts seemed to be related to students' opportunity to choose the way they could work and to overcome challenges throughout the project. The “engaging” nature of the project was demonstrated in the video data (27% of the observed incidents concerned this dimension) as well as students' statements in the focus-group interview data. Relevant findings, that is, making-&-tinkering activities influencing students' attitudes towards STEAM, have been reported in recent studies (Chu et al., 2017; Harnett et al., 2015).

*RQ3: How this approach seems to enact the development of 21<sup>st</sup>-century skills?*

The study further supports that making-&-tinkering projects can support the development of students' **21<sup>st</sup>-century skills**, particularly in all the three categories of the framework by the Partnership for 21st Century Learning (P21, 2015). Throughout this experience we documented multiple episodes of critical thinking, problem solving, creativity and

innovation, communication and collaboration. Similar results were found in previous works (Bevan, Gutwill, Petrich & Wilkinson, 2015; Harnett, Tretter, & Philipp, 2015).

Although the study, confirms to a large extent, findings already reported in previous works, the field is still in maturing and therefore replicability helps to enhance the validity of such methods and practices in education. The present study is one of very few with young children of this age (8 to 9 years old), engaging in making-&-tinkering activities in the context of formal mathematics curriculum and therefore, has a unique merit in the research literature with implications for researchers and practitioners in the area. The small (N=18) sample of convenience in the present investigation does not allow for generalizability of findings, yet the rich description of our methods and procedures should enable other researchers and practitioners to transfer these findings into similar contexts and settings. Overall, a better understanding of how making-&-tinkering occurs in formal settings with young learners will enable researchers and practitioners to nourish learning environments which can serve the development of the 21<sup>st</sup>-century skills.

## Acknowledgments

This work is part of a project that has received funding from the European's Horizon 2020 research and innovation programme under grant agreement N°739578 and the government of the Republic of Cyprus through the Directorate General for European Programmes, Coordination and Development.

## References

- Attride-Stirling, J. (2001). Thematic networks: an analytic tool for qualitative research. *Qualitative research*, 1(3), 385-405.
- Bevan, B., Gutwill, J. P., Petrich, M., & Wilkinson, K. (2015). Learning through stem-rich tinkering: Findings from a jointly negotiated research project taken up in practice. *Science Education*, 99(1), 98-120.
- Chapman, D. W., & Carter, J. F. (1979). Translation procedures for the cross-cultural use of measurement instruments. *Educational Evaluation and Policy Analysis*, 1(3), 71-76.
- Chu, S. L., Angello, G., Saenz, M., & Quek, F. (2017). Fun in Making: Understanding the experience of fun and learning through curriculum-based Making in the elementary school classroom. *Entertainment Computing*, 18, 31-40.
- Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011, September). From game design elements to gamefulness: defining gamification. In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments* (pp. 9-15). ACM.
- Ioannou, A. (2018). A model of gameful design for learning using interactive tabletops: enactment and evaluation in the socio-emotional education classroom. *Educational Technology Research and Development*, 1-26.
- Flanagan, J. C. (1954). The critical incident technique. *Psychological bulletin*, 51(4), 327.
- Gentry, M., & Gable, R. K. (2001). From the Student's Perspective—My Class Activities: An Instrument for Use in Research and Evaluation. *Journal for the Education of the Gifted*, 24(4), 322-343.
- Harnett, C. K., Tretter, T. R., & Philipp, S. B. (2014, October). Hackerspaces and engineering education. In *Frontiers in Education Conference (FIE), 2014 IEEE* (pp. 1-8). IEEE.
- Hsu, Y. C., Baldwin, S., & Ching, Y. H. (2017). Learning through making and maker education. *TechTrends*, 61(6), 589-594.
- Krieger, S., Allen, M., & Rawn, C. (2015, February). Are females disinclined to tinker in computer science? In *Proceedings of the 46th ACM Technical Symposium on Computer Science Education* (pp. 102-107). ACM.
- Martinez, S. L., & Stager, G. (2013). *Invent to learn: Making, tinkering, and engineering in the classroom* (pp. 31-41). Torrance, CA: Constructing modern knowledge press.
- Papavlasopoulou, S., Giannakos, M. N., & Jaccheri, L. (2017). Empirical studies on the Maker Movement, a promising approach to learning: A literature review. *Entertainment Computing*, 18, 57-78.
- P21- Partnership for 21st Century Learning. (2015). P21 Partnership for 21st Century Learning. Partnership for 21st Century Learning, 9. Retrieved from [http://www.p21.org/documents/P21\\_Framework\\_Definitions.pdf](http://www.p21.org/documents/P21_Framework_Definitions.pdf)
- Rode, J. A., Weibert, A., Marshall, A., Aal, K., von Rekowski, T., El Mimouni, H., & Booker, J. (2015, September). From computational thinking to computational making. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 239-250). ACM.
- Sheridan, K., Halverson, E. R., Litts, B., Brahm, L., Jacobs-Priebe, L., & Owens, T. (2014). Learning in the making: A comparative case study of three makerspaces. *Harvard Educational Review*, 84(4), 505-531.
- Vossoughi, S., & Bevan, B. (2014). Making and tinkering: A review of the literature. *National Research Council Committee on Out of School Time STEM*, 1-55.

# Parent and Educator Perspectives on the Benefits of an Online Space to Promote Offline Program Collaborative Learning

Susan K Walker, Cheuk Hei Cheng, Alize Rattenni, Caroline Reynolds, and Sam Lapham  
skwalker@umn.edu, cheng539@umn.edu, arattenn@umn.edu, reyno574@umn.edu, lapha018@umn.edu  
University of Minnesota

**Abstract.** This multiple perspective, qualitative investigation with parents and teachers/teaching assistants provided insight on values of a virtual platform to aid parent collaborative learning in a dual focus early childhood program. Benefits were viewed as complementary or supplementary to face-to-face program components and context. Parent access to information about their children not available to them during program participation may deepen understanding and strengthen feelings of connectedness, particularly important to parent-teacher relationships. While there was synergy on most platform benefits, teachers did not observe parents' value in maintaining the program climate of trust and safety, and extending feelings of community. These may be essential for the platform to represent the learning supports aligned with the offline program experience. Study findings lead to hypotheses on the role of an online environment in parents' mental models of programs for collaborative learning, and in social presence that strengthen collaborative learning.

Virtual environments and interactive digital applications hold powerful potential for extending learning beyond face to face settings. Efforts to hybridize formal K-16 education with complementary technologies show positive impacts on student collaborative learning outcomes (e.g., Heflin, Shewmaker, Nguyen, 2017). Less well studied is technology-enhanced learning in nonformal education and with adult learners, specifically parents of young children who participate in community-based education programs. These learners and settings are critically important to the wider study of computer-supported collaborative learning. There is obvious societal value to promoting parenting role competence. And nonformal education group-based programs for parents prove successful at facilitating domain and practice learning through collaborative inquiry and reflection (Campbell & Palm, 2018). The conscientious integration of technology in ways that complement nonformal program aims for parent learning is rare (Strickroth & Pinkwart, 2013). The current study extends the authors' work employing design-based implementation research (DBIR) (Penuel, Fishman, Cheng & Sabelli, 2011) to explore technology integration in a dual focus early learning program (Walker, 2017).

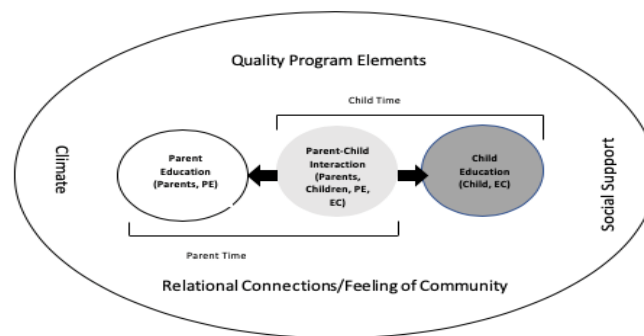
## Theoretical and empirical background

Socioconstructivist perspectives on parent learning view the acquisition of knowledge about parenting and the self in the parenting role as a continuous process, involving reflection of experience and observation of and interaction with others who reinforce norms and validate experiences (Azar, 2003). As facilitated in group parenting education, collaborative inquiry occurs in discussion, critical self-reflection and exploration of the role through the sharing of experiences. This encourages the development of mental models, or conceptual schemes that guide parent action (Lam & Kwong, 2012; Marineau & Segal, 2006). Facilitators foster group cohesion and social capital through a climate of respect, trust and security (Campbell & Palm, 2018).

Research on technology's role in parent social learning has largely focused on individual participation in what Henri and Pudlelko (2003) label 'communities of interest' (e.g., specialized groups on social media) (Nieuwboer, Fukkink, & Hermanns, 2013; Zero to Three, 2016). Parents seek others online for information and support in ways that complement the personal and professional sources in their lives (e.g., family, friends, the pediatrician) (Zero to Three, 2016). And their investment in these online contacts varies with their technology comfort, and access. To a lesser extent, parents may participate in structured online parent 'learner communities' and research suggests benefits similar to offline collaborative settings (Gray, 2004). Online formats may also blend effectively to meet offline learning program aims. For example, Guldberg and Pilkington (2006) observed that, as facilitated by a trained peer as part of a credential program, a complementary online community for the caregivers fostered context (creation of a safe space, defining common values) and process (emergence of challenging questions, resolution-focused goal attainment) dimensions of collaborative learning. Theoretically then, if parents in a nonformal education program benefit from a collaborative, constructivist format offline and experience an online complementary experience, learning gains would be greater. This was the premise for our platform design and implementation research, as discussed below.

## Social learning technology and the early childhood family education program

Minnesota's Early Childhood Family Education (ECFE) is a parenting education program that adheres to constructivist principles. Operating through local school districts, ECFE offers weekly no/low cost two-hour classes and access to community resources to parents of children from birth through age five (ECFE.info). During the classes, parents and children interact together for the first hour, guided by licensed early childhood and parenting educators, then separate for the second (Figure 1). Unlike other group parent education, ECFE does not use an established curriculum; educators tailor learning content to individual groups, guided by field guidelines of early childhood program quality (Minnesota Department of Education, 2001). ECFE sites generally offer multiple class sections per site and hold several site-wide, all-family events per year. ECFE research has demonstrated processes and contexts of learning and instruction consistent with collaborative learning principles (Campbell & Palm, 2018). Weekly discussion time with other parents, as facilitated by parenting educators, fosters learning through information exchange, reflection on experiences and joint problem solving. Attention is given to a climate and context of support and respect. The first authors' own research indicates that parent relationships fostered in ECFE endure beyond the program and are supports throughout their child's life (Walker, 2018b).



**Figure 1.** Elements and Program Structure of ECFE. (PE = parent educator; EC = early childhood educator).

Partnership with an ECFE site in an urban school district resulted in the development of an online platform to support ECFE's learning objectives (Walker, 2017). Participatory design research revealed parent preference for between-class social connectivity, information about parenting and about their child's learning; employing technology that was easy to use, mobile and ensured privacy (e.g., restricted access to class pages). The developed platform features a Facebook-like news feed for posts and discussion, and area for announcements, a shared calendar, dynamic photo album, external links, and members visual display (Figure 2). A user dashboard provides access to class pages, private messaging and notifications. Posts can be directed to the full site or to selected classes. Launch of the completed platform occurred in early 2017. Google Analytics tracking to the end of the school year revealed steady use by repeated visitors, with approximately 2/3 of sessions (i.e., group of site interactions per visit) by parents. Expanded platform use by additional ECFE sites in the subsequent school year showed similar proportionate use by parents (Walker, 2018a).

To date, investigations of the virtual platform show promise regarding usability, perceived usefulness by parents and staff and integration into program operations (Walker, 2018a). Contributing to new learning theory (Penuel, et al, 2011), requires deeper examination into cross-environment collaborative learning by parents inclusive of the roles played by teachers. Parenting educators are key to the parents' meaning making process (Lam & Kwong, 2012). And in early childhood settings that also serve parents, the perspectives of teachers and parents represent their adult roles (teacher; learner, parent) and are proxy for the children who also participate (Katz, 1993). Multiple perspective research allows a more complex and nuanced understanding of the shared experience [so that] views can be situated within the social relationships wherein they are constructed (Vogl, Zartler, Schmidt & Rieder, 2018, p.179).

## Method

In-depth interviews with multiple perspective qualitative analysis were employed to understand shared and unique perspectives from teachers, teaching aides, and parents on technology's role in promoting parent collaborative learning in ECFE.



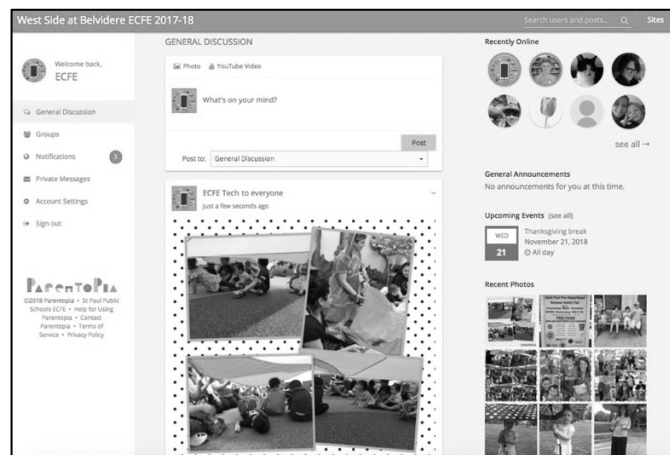


Figure 2. Desktop view of ECCE virtual platform.

## Sample

The sample consisted of 25 parents (19 women, six men) and six ECCE staff members (five women, one man) from the original ECCE design site. That site and parents for the study were selected for their demographic diversity and representativeness of families who attend ECCE in the school district. Parents were also sampled to represent all 8 weekly site classes. They ranged in age from 20 to 57 years ( $M = 35.3$ ,  $S.D. = 1.5$ ), and reported from one to six children ( $M = 2.7$ ,  $S.D. = 1.7$ ). Five parents reported caring for children with disabilities. Six parents came from the two Spanish-speaking classes. Parents were fairly evenly distributed by length of program participation (first year to more than 3 years). Staff held positions as licensed teachers (2 parenting educators, 1 early childhood teacher; all female) and as classroom assistants (3; 1 male). Their years of working in ECCE ranged from three to 35 years ( $M = 21.3$ ). Staff ages ranged from 39 to 56 years ( $M = 48$ ).

## Data collection and analysis

Semi-structured interviews explored participation or instructional experiences with the ECCE program and the use of technology in ECCE. Each interview lasted on average 1 hour (range: 45 minutes to 3 hours, 30 minutes). Audio-recorded interviews were conducted at the ECCE site by one of the three trained interviewers. Informed consent was obtained from participants prior interviews. Interviews were transcribed to text, then audited for accuracy by a second transcriber, and pseudonyms applied to all names identified for confidentiality.

An informed constructivist grounded theory approach (Charmaz, 2014; Thornburg, 2012) guided coding and analysis. Additional considerations in coding and analysis were deployed to compare within and across groups consistent with multiple perspective analysis (Vogl, et al 2018). The coding process with memoing consisted of three phases, initial coding, focused coding, and theoretical coding. Coders were assigned in pairs to analyze parent and staff interviews separately and together to enhance trustworthiness (Denzin, & Lincoln, 2018). During initial coding, transcripts were read and coded with the incident-by-incident approach, informed by sensitizing concepts from the literature on collaborative learning in parenting education. Second, focused codes emerged after reviewing initial codes. Themes identified within one relational unit (e.g. parents, teaching staff) were compared by coding pairs to generate central theoretical codes by unit. Finally, theoretical codes emerging from each relational unit were compared with each other. The themes emerged from staff and parents were compared across pairs of coders to understand the meanings of the perspectives, and shared and differing aspects of the perspectives.

## Results

A summary of parent and staff perspectives on collaborative learning as supported by the virtual platform is provided in Table 1. Learning supports were identified as process (e.g., aiding the acquisition or interpretation of content) or context (e.g., climate, tone) and as being supplementary (in addition to) or complementary (an extension of) the face-to-face program experience.

Table 1. Perspectives on technology benefits to collaborative learning in ECCE

| Construct   | Evidence  |
|---|---|
| Information about Children<br><i>Process/Supplementary</i>  | <p>Parents</p> <ul style="list-style-type: none"> <li>● Positive emotion at seeing pictures of children in classroom while away</li> <li>● Feel taken care of, thought of by staff</li> <li>● Understand more about child and children</li> </ul> <p>Staff</p> <ul style="list-style-type: none"> <li>● Opportunity to add information that parents don't get while in parenting education</li> <li>● Reflect interest by parents to learn more about children in classroom</li> <li>● Strengthen relationships between parents and teachers:</li> </ul>  |
| Engagement with parenting content<br><i>Process/Complementary (main)/ Supplementary (secondary)</i> | <p>Parents</p> <ul style="list-style-type: none"> <li>● Access information if can't attend (supplementary)</li> <li>● Share information with co-parent who can't attend (supplementary)</li> <li>● Gather information from parents in other class supplementary)</li> <li>● Read or view before or after class (complementary)</li> <li>● Get advice and information from parents and teachers (complementary)</li> </ul> <p>Staff</p> <ul style="list-style-type: none"> <li>● Content to stimulate thinking before or after class (complementary)</li> <li>● Additional materials provided in handouts, drive links, URLs to deepen learning (complementary/supplementary)</li> </ul> |
| Relational/supportive engagement with families<br><i>Process/Complementary</i>                      | <p>Parents</p> <ul style="list-style-type: none"> <li>● Support to parents from other parents, from staff</li> <li>● May use with other methods to connect with family and friends for support</li> </ul> <p>Staff</p> <ul style="list-style-type: none"> <li>● Continue to provide 1-1 communication of support to parents.</li> <li>● Sharing images and videos with other children and families helps retain connections beyond class, over summer.</li> </ul>   |
| Program communication and information<br><i>Process/Complementary</i>                               | <p>Parents</p> <ul style="list-style-type: none"> <li>● Easy access to program reminders, events, 1-1 communication</li> <li>● Maintains 'idea' of program continually availability</li> </ul> <p>Staff</p> <ul style="list-style-type: none"> <li>● Easy provision of program reminders, events, 1-1 communication</li> <li>● Makes work more efficient; easier to complete administrative tasks</li> </ul>  |
| Climate, tone, safety, privacy and trust<br><i>Context/Complementary</i>                            | <p>Parents</p> <ul style="list-style-type: none"> <li>● Feels like a safe space to post and view pictures of my child</li> <li>● Tone in discussion similar to what is experienced face to face</li> <li>● Appreciate member only access for privacy</li> </ul>   |
| Maintain Community<br><i>Context/Complementary</i>  | <p>Parents</p> <ul style="list-style-type: none"> <li>● Able to maintain connections to others that extends community</li> <li>● Able to connect and see families who attend other classes. Represent diversity in community.</li> </ul>  |

### Supplementary information access: Child

Of all identified, this was the dominant subtheme expressed by parents and by staff. Both observed how teachers and occasionally, other parents posted pictures or videos of children. At times they augmented the images with text to identify developmental gains or processes. Parents expressed appreciation for the chance to learn more about their child, and what their child was doing in the classroom when they were in parenting education. Parents spoke with pride at seeing their child's activity or accomplishment; some also spoke of learning more about the child in context with other children. They also felt cared for by staff. Mother of 2, Mia said, "So it's nice to, like, have them vested in, like, 'hey the parents probably want to see their kids doing this'." For early

childhood staff who have limited interaction time with parents, it was the opportunity to help parents learn more about the child and child development. Staff also shared positive emotions about the exchange: *Teacher Jade: "You know, I think that, that's a priceless tool, because in the moment I can have that [image of what a child is doing] with me, and I can look back to it; I can share it with the parent. I can say that I'm noticing something changing. We can look for patterns together.* For both parents and early childhood staff, the opportunity and action in posting and reading/seeing information about the child appeared to strengthen the sense of connection.

### Supplementary information access: Parenting

Parent access to information about parenting when they weren't in class due to absence or inability to attend (as is the case of many co-parents) was less frequently but consistently mentioned across parents and the two parenting educators. The parenting educators felt that this helped them better reach all members of the class learning community. Parents appreciated feeling caught up. Supplementary access also meant learning from someone in a different class. Stella noted, *"She had a special needs child ... and she put an article online that we could all read.... that's very cool and I think we need to know more about that...[there are] cool things that I can access now and I can learn more and I can become a better parent now because I get to see what everybody's doing"*

### Parenting information complementary to classroom discussions

Parenting educators appreciated being able to post a video, reading or webpage to stimulate thinking before the weekly class or extend discussion afterward. Although parents were generally positive about having access to this information (e.g. *"I think that would be really valuable because there's stuff that we're all interested in and I'm sure there's information and resources that it's being drawn from and it might be interesting to spend you know rather than flipping through Facebook, one night."*), few spoke about continuing discussions outside of class. As mother of 6 Bella observed,

*"You know and that's kind of sad but it's the truth because our lives are so busy and fast moving. And then for me, I don't want to scroll through and see like what questions someone asked today, because they're not really my questions. I don't know. So if I have a burning question or something that's really is on my mind, I'm not going to probably take the time to like post it and then wait to see if anyone responds. Because it's not immediate as a text would be or a discussion with someone face to face."*

### Continued access to program information

Parents and teachers liked easy access to information for registration, events, activities and reminders. The platform was characterized as a one stop location to coordinate and find site information. Parents remarked that this helped simplify their busy lives; staff liked the efficiency of reaching more parents, more quickly. Reduced time in email and phoning, freed up time for teaching or other program responsibilities.

### Access to support from parents and teachers

Use of the platform as a vehicle to continue supportive connections between parents, and from staff to parents was a common theme across parents and teachers, especially the parenting educators and one of the early childhood aides (who'd worked in ECFE for over 20 years). Three parents from one class spoke of using the platform to organize meals for an ECFE family who'd just given birth to triplets. Parents also relayed a sense of relief or gratitude to extend 1-1 access to the parenting educator. In turn, staff commented that the platform complemented other virtual and real time means for providing support to parents. Parenting educator Everly said: *[...] you know I give parents my phone for texting, e-mail, [the platform] is great for that of course. And parents will call me, parents will text me. They'll reach out and...and not just me. I'm saying that sometimes it's a starter place. But I encourage them also to reach out to each other. So that's where the platform is helpful for that as well."*

### Continuity of community and climate (parents)

A difference between parents and staff related to the platform extension of feelings of community, and program climate and tone. Parents observed this; staff did not. This quote from 21 year old Brooklyn conveys the extension of community:

*"And just being able to kind of like someone's picture of their kids you know kind of makes you feel like you're friends with that person and so it has that very Facebooky feel to it and that*

*creates I think a kind of that community. And maybe give you an idea, you know, 'oh I saw that picture of you guys doing this and that looks fun' and it gives you maybe something to talk about when you see them in class...I have used it to connect with people outside of class and say like 'hey let's have a playdate go to the park' kind of thing. So that's a nice way to keep building that community."*

Parents were also vocal about the maintained tone of civility and the sense of safety that they felt from the private, ECFE member only access. As Lily stated:

*"...having like a respectful place that we had just communicate with each other that you still feel like, ECFE feels like a safe space to me and I feel like Parentopia is kind of like, also it offers that as well, because they know that there is an administrator and there's not going to be any kind of negative thoughts or comments. I mean everyone's really supportive. Yeah and I think it's a place that I can kind of share my ideas too and with everyone and they can do the same"*

## **Conclusion and implications**

This study offers preliminary support that parents and staff alike find shared and unique values to an online platform as an extension of the face to face weekly program, and identify a variety of benefits to collaborative learning. Platform interactions allow for continuity of connection (important for parents with infants, multiple children and single parents, and for co-parents and other caregivers who cannot attend). It extends access to parenting content, emotional and informational support, and to program events and activities. And significantly, it serves as a means to receive information about the child that the parent can't acquire due to the program structure that separates parents and children. Staff see the platform as improving the quality of their teaching and outreach, it may make their work easier and more efficient, and this may bring a sense of emotional satisfaction. While there is potential evidence of the platform's practical value to collaborative learning, insights from this study also suggest possible theoretical or conceptual advancements.

## **Strengthening parent-teacher partnerships and educator presence**

In the traditional ECFE program, parents separate from their 2 ½ -5 year old children during parenting education time and leave their children in the capable hands of the early childhood staff. While they look forward to time to themselves and the interaction with other adults for reflection, sharing and discussion, many wonder about their children's activity and happiness. The platform allows staff to post information, videos and text-augmented photos of the child(ren). Because of the community (class) and site (families) nature of the program, pictures of the child with other familiar children can help reinforce the parent's sense of the child in a wider social context. Parents expressed a range of positive emotions when voicing appreciation for the information about their child provided by the staff. Many conveyed sentiments of feeling closer to staff. In turn, early childhood staff who cannot interact with parents as much during the all family parent-child interaction time appreciate the ability to share information in other ways. Parents' sense of learning community membership is not exclusive to peers, but includes their children and ECFE staff (Walker, 2018b). Children are the reason that parents join ECFE, and raising healthy children is the value that all adults – parents and teachers - share in the program. Supplementary interactions may strengthen the sense of trust and connection between parents and professional staff who teach and care for the child (Lauluvein, 2010). Stronger parent-teacher relationships in early childhood bode well for children's learning outcomes and set the stage for partnerships in later school years (Froiland, Peterson & Davison, 2012).

This aspect of collaborative learning by parents facilitated online by teachers' actions may also have an emotional component. Sharing and receiving information about the child at the center of the teacher and parent's interest in ways that bring satisfaction to both may deepen the learning experience and offer a reciprocal interaction that sustains learning online and offline. Teachers like Everly observed the benefits to relationships: *"the more the parents see the teachers posting and they see their name up there and they see their face and they see what they posted they see a photo. Parents are bridging- they're connecting more. They're building stronger relationships."* Lehman's (2006) framework of presence in online learning suggests that addressing emotional considerations in interaction activities can help learners' confidence, focus and excitement to learn collaboratively.

## **The importance of community and climate: Lessons for professionals**

Whereas parents appreciated how the platform mimics the climate, tone, privacy and feelings of safety of the face to face site, and extended community and connection to others who share a parenting and program identity, staff didn't speak to these. The difference may be due to the 4 to 1 ratio of parents to professionals in the study, affording more observations by the former. The difference however, may be due to differences in perspectives of *being* a community member vs one whose *job* it is to foster learning as community. Staff were more articulate about the specific technologies they use for instruction and communication, and pedagogical purposes related to sharing information, communication, extending class content and lending support. A focus on technology for teaching over the social experience of the learner may limit teachers from their seeing themselves as participants in discourse or as members of parents' learning community (Lauluvein, 2010). They may focus on providing content when a busy parent with other content resources may more highly value their presence and affirmation.

Staff may need guidance to extend community in ways that maintain the visual membership and emotional tone of the face to face program. This means encouraging families to post information about themselves, pictures from neighborhood events, and posting in a site-wide forum. It can also mean monitoring discourse to maintain exchanges in the civil, positive and nonjudgmental tone that led parents in this study to make favorable comparisons to the platform over their less than positive experiences with Facebook. Encouraging a respectful online climate and sense of community can help parents make connections across classes and build bridging social capital valuable to learning (Wenger, White & Smith, 2009).

### The platform as representation of whole program participation experience

Parents, teachers and teaching assistants identified a range of ways in which using the platform served to complement the weekly face to face learning experience. These included exposure to parenting content curated by professionals, provision of up to date program information, access to supportive others, outlets for ongoing communication to peers and professionals, and continued visual and text reminders of membership in a learning community. As these informational, interactive and relational elements align with constructivist, collaborative inquiry approaches in parenting education (Campbell & Palm, 2018; Lam & Kwong, 2012) and represent the values and operation of ECFE, they may indicate the potential of a virtual platform as useful to the multiple process and context mechanisms that foster parent collaborative learning.

Theoretically however, the platform may also facilitate parent learning from a conceptual standpoint. That the platform serves many learning and engagement functions complementary to parents' ECFE experience may support parents' mental image of themselves as participating in a learning community beyond the weekly session structure. The notion is expanded as they are also "in the children's classroom" through the supplementary images and information about their children. This may give them a sense of participating in all components of the weekly program depicted in Figure 1. And as ECFE is a place that is positive, as busy, stressed parents, the platform sends the message of ECFE being continually present, reliable and responsive, and this may bring them comfort. This includes the consistently strong presence of staff who for some, function as significant emotional supports. So, the platform's ability to extend the 'idea' or mental model the program that parents affiliate with as a supportive space for learning with others, including staff, may be another complementary value of the platform for parents' collaborative learning.

This study is, to be sure, preliminary in its findings. It depended on the voices of parents and staff at a single ECFE site, about their experiences with a newly designed platform early in its implementation. Yet, there is promise in the findings to buoy continued platform implementation in the school district program that will allow use by larger numbers of parents and staff, at more ECFE sites. And it encourages further investigations in the program-platform coordinated role in parent learning. To date the platform is being used by eight ECFE programs in four school districts and reaching over 450 parents and teachers (Walker 2018a). The current findings validate employing design-based implementation research when identifying innovation for a long-standing face-to-face parenting education program. Attention to the needs of the program and of the participants and staff allowed development of a useful and possibly critical innovation to face to face parent education and extending adult learning benefits. Continued research on implementation of the platform will suggest avenues for its contribution to learner success and for sustainability within nonformal, community-based parent education settings.

### References

- Azar, S. T. (2003). Adult development and parenthood: A social-cognitive perspective. In J. Demick & C. Andreoletti (Eds.), *Handbook of adult development* (pp. 391-415). New York: Springer.
- Campbell, D. and Palm, G. (2018). *Parent education: Working with groups and individuals* (1st ed.). San Francisco, CA: Jossey-Bass.
- Charmaz, K. (2014). *Constructing grounded theory*. London; Thousand Oaks, Calif.: Sage.

- Denzin, N. K. and Lincoln, Y. S., Editors. (2018). *The SAGE handbook of qualitative research*, 5th edition. Thousand Oaks, CA: Sage.
- Froiland, J. M., Peterson, A., & Davison, M. L. (2012). The long-term effects of early parent involvement and parent expectation in the USA. *School Psychology International*, 34(1), 33-50. doi: 10.1177/0143034312454361
- Gray, B. (2004). Informal learning in an online community of practice. *Journal of Distance Education*, 19, No 1, 20-35.
- Guldborg, K. & Pilkington, (2006). A community of practice approach to the development of non-traditional learners through networked learning. *Journal of Computer Assisted Learning* 22, 159–171.
- Heflin, H., Shewmaker, J., & Nguyen, J. (2017). Impact of mobile technology on student attitudes, engagement, and learning. *Computers & Education*, 107, 91-99.
- Henri, F. and Pudelko, B. (2003). Understanding and analyzing activity and learning in virtual communities. *Journal of Computer Assisted Learning*, 19, 474-487.
- Katz, L. G. (1993). Multiple perspectives on the quality of early childhood programs. *European Early Childhood Education Research Journal*, 1(2), 5-9.
- Lam, C. M., & Kwong, W. M. (2012). The “paradox of empowerment” in parent education: a reflexive examination of Parents' pedagogical expectations. *Family Relations*, 61(1), 65-74.
- Laluvein, J. (2010). Parents, Teachers and the "Community of Practice". *The Qualitative Report*, 15(1), 176-195.
- Lehman, R. (2006). The role of emotion in creating instructor and learner presence in the distance education experience. *Journal of Cognitive Affective Learning*, 2(2), 12-26.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry* (Vol. 75). Sage.
- Marienu, C. & Segal, J. (2006). Parents as developing adult learners. *Child welfare*, 85(5), 768-784.
- Minnesota Department of Education (2001). Early childhood quality indicators of progress. St Paul, MN. <https://education.mn.gov/MDE/dse/early/ind/>
- Nieuwboer, C. C., Fukkink, R. G., & Hermanns, J. M. (2013). Online programs as tools to improve parenting: A meta-analytic review. *Children and Youth Services Review*, 35(11), 1823-1829.
- Penuel, W. R., Fishman, B. J., Cheng, B., & Sabelli, N. (2011). Developing the area of design-based implementation research. *Educational Researcher*. Menlo Park, CA: SRI International.
- Strickroth, S., & Pinkwart, N. (2013). An Approach for Supporting Hybrid Learning Communities: The Case of a Regional Parent Community. In *Proceedings of the CSCL*.
- Thornberg, R. (2012). Informed grounded theory. *Scandinavian Journal of Educational Research*, (56), 3, 243-259.
- Vogl, S., Zartler, U., Schmidt, E. M., & Rieder, I. (2018). Developing an analytical framework for multiple perspective, qualitative longitudinal interviews (MPQLI). *International Journal of Social Research Methodology*, 21(2), 177-190.
- Walker, S. (2018a). Parentopia: Overview of social interaction technology to extend family engagement and support. University of Minnesota. [https://drive.google.com/file/d/1D\\_ee7Jp8\\_Y70tvqXeHYVKGeUFVnRfeuD/view?usp=sharing](https://drive.google.com/file/d/1D_ee7Jp8_Y70tvqXeHYVKGeUFVnRfeuD/view?usp=sharing)
- Walker, S. (2018b). I think it's made everyone a little closer: A virtual platform extending parent learning as Community of Practice. Poster presented at the 2018 International Society for Learning Sciences conference, London, June 29.
- Walker, S. (2017). Creating Parentopia: Design-based Research to Develop an Interface for Parent Learning Communities and Networks. Computer Supported Collaborative Learning bi-annual conference. Philadelphia, PA
- Wenger, E., White, N. and Smith, J.D. (2009). *Digital habitats: Digital habitats: Stewarding technology for communities*. Portland, OR: CPSquare.
- Zero to Three (2016). *Tuning In: Parents of Young Children Speak Up About What They Think, Know and Need* Washington, DC. <https://www.zerotothree.org/resources/>

# An Alternate Statistical Lens to Look at Collaboration Data: Extreme Value Theory

Kshitij Sharma, Department of Computer Science, Norwegian University of Science and Technology,  
Kshitij.sharma@ntnu.no

Jennifer Olsen, Computer Human Interaction Lab in Learning and Instruction, École Polytechnique Fédérale de  
Lausanne, Jennifer.olsen@epfl.ch

**Abstract:** To provide beneficial feedback to students during their collaboration, it is important to identify behaviors that are indicative of good collaboration. However, in a collaborative learning session, students engage in a range of behaviors and it can be difficult to indicate which of those behaviors correlate with higher outcomes. In this paper, we propose using Extreme Value Theory (EVT), a method that considers the data points in the tail (upper or lower) of the distribution, to analyse the relationship between collaborative process variables and outcome measures through insights derived from high impact, low-frequency events. Specifically, in this paper, we analyse the relationship between dual gaze patterns and outcome measures across two different datasets. In both datasets we found that students with lower outcomes had lower focus during the collaborative session. This paper provides a contribution by both introducing EVT as a viable method for analysing CSCL data as well as demonstrating the effectiveness of eye-tracking as a collaborative indicator to use to adapt to in real-time.

**Keywords:** Eye-tracking, dual eye-tracking, extreme value theory, CSCL, collaborative process, intelligent tutoring systems, concept maps.

## Introduction

Collaboration can be an effective way of learning, but it is challenging to ascertain how students' actions lead to learning when working in a group and how to determine what actions should be considered when providing students with feedback on their collaboration. To provide timely and effective feedback, it is important to have an indicator of collaboration that can easily be measured in real-time and is a predictor of learning. However, across collaboration indicators, the connections between these student process measures and learning outcomes have not always been clear (Deiglmayr, Rummel, & Loibl, 2015; Olsen, Alevén, & Rummel, 2017). In this paper, we propose analysing only the high impact, low frequency data points to provide additional insights into the relationship between process data and collaborative learning outcomes. We introduce the use of Extreme Value Theory (EVT), a method that considers the data points in the tail (upper or lower) of the distribution, to analyse the relationship between collaborative process variables and outcome measures. Specifically, we are interested in dual eye-tracking as a process variable as it can easily be measured in real-time and be used to provide students with feedback (Sharma et.al, 2016, D'Angelo & Begel, 2017). Using EVT, we aim to increase our understanding of the process variables that impact student collaborations and how they can be used to provide students with real-time feedback.

EVT is a novel method to CSCL and presents a complementary viewpoint to the statistical methods often used to analyse collaborative learning data while having fewer data assumptions. EVT can be used for both explanatory analyses (Ramesh and Davison, 2002) and hypothesis verification (Santinelli et.al., 2014). The mathematical foundations of EVT are similar in strength, rigor and maturity as compared to the methods based on the central tendencies in the data (Smith 1990). EVT provides a way to estimate the probability of occurrence of rare events, which might also be unseen in the observed data. This makes EVT a unique method to implement proactive feedback to support collaboration. For example, in the context of collaborative eye-tracking, if we could estimate the probability of peers not focusing on any part of the communication mediating interface, we could suggest some remedial actions in a proactive manner. In contexts where the proactive feedback is required, the tail of the data may contain more information about the process than the main body of the distribution.

In this paper, we apply EVT to eye-tracking as it has become a key source of process data in educational research over the past few years. Research using eye-tracking covers a wide range of educational ecosystems from online (Sharma et. al., 2015) to face to face classes (Raca & Dillenbourg, 2013), from co-located (Schneider et. al, 2016) to remote collaborative learning (Sharma et. al., 2012), and to understand teachers' classroom orchestration processes (Prieto et. al., 2016). Eye-tracking has not only been used to understand the learning processes in various contexts, but it also has been used to provide students appropriate, real-time, and adaptive feedback on their learning processes (Sharma et.al, 2016, D'Angelo & Begel, 2017).

One of the most common practical uses of dual eye-tracking (DUET) data in CSCL has been to quantify collaborative outcome and processes. In terms of collaborative outcome, recent results have shown DUET measures of gaze cross-recurrence (looking at the same area of the screen at the same time) to be useful in quantifying collaboration quality (Schneider et. al., 2013; Jermann and Nuessli, 2012) and the differences between the expertise (Papavlasopoulou et. al., 2017). Further looking at the similar areas of the screen in each time window was found to be correlated with learning gains (Sharma et. al., 2015; Schneider et. al., 2016; Sangin et. al., 2011). DUET measures such as average entropy (Sharma et. al., 2012), transitions among areas of interest (Villamor and Rodrigo, 2018) were correlated with task based performance in pair programming tasks. In terms of collaborative processes, eye-tracking research has shown to be useful in relating gaze to dialogue patterns such as the eye-voice span (time difference between the speaker's gaze at an object and verbalization, Allopenna et al., 1998), voice-eye span (time difference between the speaker's voice and listener's gaze to the referred object, Griffin and Bock, 2000) and eye-eye span (time difference between the speaker's gaze at an object and the listener's gaze at the same object, Richardson et al., 2007). Gaze cross-recurrence was also useful in explaining the conceptual knowledge gains of the peers in a collaborative learning scenario with intelligent tutoring systems (Belenky et. al., 2014). One common theme across these studies are the statistical methods used. Most of these studies used methods based on the central tendencies of the data. We propose a complementary viewpoint for the DUET data in this contribution that is based on the extreme values present in the data.

Within this paper, we aim to answer the research question of how the visual focus of peers is related to the collaborative learning outcomes. We propose two types of shifts from the traditional analyses. First, we move from the main body of the distribution (central tendency) to the tails of the distribution (extremes). Second, we move from the individuals (or groups) as a unit of analyses to the specific moments in the interaction. Specifically, we are interested in *how the extreme values (i.e., the moments with extremely high or extremely low values) can inform us of the students' learning gains*. To answer this question, we present the EVT method in two different dual eye-tracking (DUET) contexts. The first study is in a collaborative concept map context with university students, and the second study is in a collaborative Intelligent Tutoring Systems (ITS) context with elementary school children. For both contexts, we hypothesized that the gaze of the students with lower test scores will have a higher tendency to wander all over the screen. In the following sections, we will present the EVT methodology and how it can be used to provide insights into effective student collaboration processes. This paper contributes to the CSCL literature by providing an alternate method for analysing relationships between process and outcome data that complements existing methodology while also extending our understanding of the relationship between visual focus and collaborative outcomes.

## Method

### Extreme Value Theory

Extreme events are those with high impact and low frequency (HILF), and EVT is the branch of statistics used for modelling HILF events. The basic idea is to model an extreme event in a way such that the analyses and implications account for the needs of unprecedented situations. For example, in finance and environmental sciences, HILF events are often analysed to allow people and companies to be proactive against negative events (e.g., risk, losses, natural disasters). We can extend this proactive policy to education by considering negative and positive HILF events. EVT deals with asymptotic data where the data used in the analyses is a small subset of the whole data -- usually below 5th percentile or above 95th percentile. Mathematically, EVT is based on the tail of a given distribution. Because EVT is based on the tails, it does not impose any assumption on the distribution and can be applied to any known (e.g., normal, student, uniform, exponential) or unknown distribution. EVT was developed initially for independent and identically distributed variables (for a complete introduction to Extreme Value Theory, see Coles, 2001), but it can easily be extended to stationary variables (variables whose distributions are not affected by the time shifts), which are most variables addressed in the CSCL community.

### Advantages and disadvantages

There are three main advantages of EVT. **First**, since EVT does not have any assumptions for the data distribution, it can be used to analyse the tail (data below the 5th percentile or above the 95th percentile) of any given distribution. This makes EVT applicable with those datasets which cannot be analysed using the common parametric approaches. For example, ANOVA requires the data to follow a normal distribution and if not, normalisation operations can affect the interpretability of results. **Second**, when analysing the dependence structure between two variables, such as correlations, EVT does not assume the dependency to follow a known structure. For example, in case of correlations, this structure is assumed to be linear. This lack of assumption allows EVT to be applied irrespective of the nature of the distribution that generates the data and dictates the



relationships among different variables. The **third** advantage of EVT is over non-parametric models. Non-parametric models, which are used in CSCL as a way of hypothesis testing, provide one value (p-value). These methods summarise the data and can handle only low dimensional problems. Given the advents in big data, non-parametric, which are designed for smaller datasets, are at an inherent disadvantage. Also in time series analyses, dynamic models provide much more information about the process than non-parametric models. EVT accounts for time series analyses by providing methods to consider the covariate dependence on time while analysing CSCL data.

The main disadvantage of using EVT is related to the fact that EVT considers the tail of the data only. By only considering the tail, the parameter estimation becomes more difficult for datasets with few events. In other words, EVT is appropriate to be used contexts where the data is sufficiently large. However, in the “big data era” this problem is often solved by itself and in these circumstances the methods based on EVT provide robust approach for estimating the probabilities of rare events.

### EVT: Univariate

There are two ways to model data using EVT: blockwise maxima and points over threshold. In **blockwise maxima**, we divide the set of observations into M blocks of N data points. This results in a sequence of maximum values for each block. One of the most important results of EVT formulation is that for any given distribution for the main body of data, known or unknown, the sequence of maximas follow a unique distribution called Generalised Extreme Value (GEV). Once we have the sequence of maximas, we simply fit the GEV distribution on them using any likelihood optimisation technique. This distribution is characterized by three parameters: position, scale and shape. These values are then used to estimate the value which has a low probability of being exceeded (see further in this section).

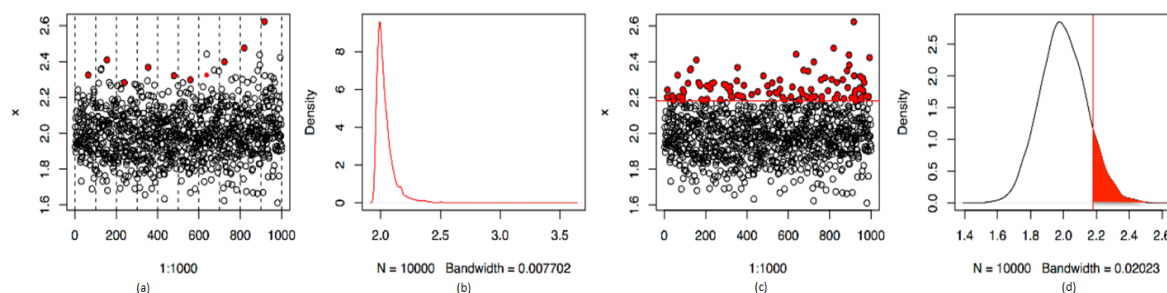
Using **points over threshold method**, we consider all the data points over a threshold. This threshold is usually a high percentile of the observed data (90th or 95th percentile). These data points above the threshold are called the exceedances. Using points over threshold method, the exceedances follow a Poisson distribution and the exceedance size follows a Generalised Pareto Distribution (GPD). The parameters of GPD can be estimated by GEV. The shape parameter for both GPD and GEV remains the same. This parameter determines how long/heavy the tail of the observed data is. Once we have the threshold, the rate parameter of the Poisson distribution and the scale and shape parameters of GPD can be estimated, like GEV, by using a likelihood optimization technique. Figure 1 shows the difference in the two approaches.

The position (or rate of Poisson), scale and shape parameters are difficult to interpret in CSCL situations, since these values do not correspond to any behavioural indicator. Therefore, we calculate a quantile value at a high level (above 90th or 95th percentile), which has an important interpretation. This value is called **Return Level**, which represents a measure of an extreme event, possibly unseen, with a certain probability. For example, if a return level is calculated at the 95th percentile, this indicates that the actual (unseen) extreme event will exceed this value with a 0.05 probability. In other words, a return value at 95th percentile is the value that has a probability of 0.05 of being exceeded. One can estimate this value by position, shape and scale measures from GEV or rate of arrival of exceedances, shape and scale measures of GPD.

One might inquire that if the high percentile can be directly calculated from the data, then why we need this estimation. The answer to this question is that we could surely calculate the return levels from the data, however minor discrepancies in the data would result in large errors for such computation, resulting in an erroneous return value larger or smaller than expected. If the return value is larger than it should be then one cannot proactively take actions; and if the return value is smaller, then one will be acting upon at a wrong time. It can be shown, mathematically that estimating GEV or GPD parameters are the correct way of computing the return level in the data. Finally, once we have the return level for the collaborating partners, we can compute the difference between them: by value and/or when they appear, and this difference can be used as a measure to further correlate against the quality/outcome of collaboration.

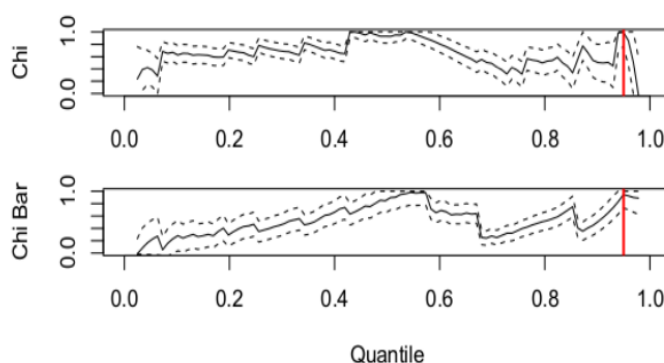
### EVT: Bivariate

To analyse the time series data from peers, the bivariate case of EVT can be useful. Bivariate EVT measures the extremal dependence between two time series data, such as that between a collaborating dyad. It models the probability of observing an extreme event in one time series given that there is an observable extreme event in the other one. This probability can be quantified using the tail-dependence between the two time series. In a classical statistical approach, the dependence between two time series is measured by correlation, which is computed using the central tendencies of the data. In the case of EVT, the tail-dependence is calculated at the high percentile of the data, as in case with the return levels.



**Figure 1.** (a) A random variable simulation with the blockwise-maxima; (b) density plots for one of the blocks; (c) the red horizontal line shows the threshold for the POT method; (d) the whole distribution, the red vertical line shows the threshold for the POT method and denotes the beginning of the tail for the distribution.

There are two measures of extremal dependence, originating from classical multivariate EVT: asymptotic dependence and asymptotic independence. The coefficient of asymptotic dependence (CAD) is the tendency for one variable to be over a high threshold when the other exceeds this threshold. This value is always between 0 and 1. The only possibility of asymptotic independence is when CAD is 0. When the CAD is greater than 0, the variables are asymptotically dependent. On the other hand, the coefficient of asymptotic independence (CAI) is the measure of strength of this extremal dependence. This is measured by a conditional probability that the smaller values in the time series of one variable are below a infinitesimal threshold, given the smaller values in the other time series are below that threshold. Mathematically, it can be shown that a value of 1 for CAI shows the perfect dependence and a value of 0 for CAI shows perfect independence. In summary, the CAD shows the level of asymptotic dependence between two time series while the CAI shows the strength of this dependence. Figure 2 shows an example of how to determine CAD and CAI.



**Figure 2.** Example illustrating the determination of the coefficient of extremal dependence and the strength of dependence for a typical time-series pair. The dashed lines represent the 95% confidence intervals for CAD (top, Chi) and CAI (bottom, Chi Bar). The tail-dependence and its strength is determined by the values at the higher quantiles (typically between 95% and 99%). The red lines correspond to 95%.

## Examples

In this section, we provide two different examples from distinct collaborative learning scenarios. For each context, we show how EVT can be applied on one gaze variable computed from both the studies, and how the different EVT based measurements differentiate the quality/performance levels in collaborative learning outcomes.

### Collaborative concept map

This data set involves 24 dyads from a larger study that tested a hypothesis about the relation between individual and collaborative gaze patterns (Sharma et. al., 2015). Each dyad was engaged in a collaborative concept-map building activity. The students were sitting on two sides of a visual separation and could talk to each other. Prior to the concept-map building activity, they individually watched two videos from Khan Academy about resting membrane potential and they were asked to build the concept-map about the same topic. The main task was to relate the pre-defined concepts and add new concepts if they felt necessary. The students watched the video at their own pace and the total duration for the concept-map activity for each dyad was between 10 and 12 minutes. For this contribution, the dependent measure is calculated as follows. The final concept-map was compared with the concept-map created by the two experts. The pair received a score using the following rules: 1) one mark for

each correct connection between two concepts, 2) one mark for each correct label of the edge between two concepts, 3) half a mark for each partially correct label of the edge between two concepts. The pairs were then divided into two levels based on the concept-map score using a median split.

## Collaborative intelligent tutoring systems

This data set involves 14 4th and 14 5th grade dyads from a larger study that tested the hypothesis about differential benefits of collaborative versus individual learning (Olsen et. al., 2014). The dyads were engaged in a problem-solving activity around fractions using a networked collaborative ITS, which allowed them to synchronously work in a shared problem space where they could see each other's actions while sitting at their own computers across the room from each other. The students could communicate verbally through a Skype connection. Each dyad worked with the tutor for 45 minutes in a pull-out study design at their school. The morning before working with the tutor and the morning after working with the tutor, students were given 25 minutes to complete a pretest or posttest individually on the computer to assess their learning. During the experiment, dual eye tracking data, dialogue data, and tutor log data in addition to the pretest and posttest measures were collected. For this contribution, the dependent measure is the average posttest score of each dyad.

## Variable: Spatial Entropy

To capture the visual focus, we use Spatial Entropy (SE) that is one of the measures used to analyse DUET data in previous research (Olsen et. al., 2018; Sharma et. al., 2018) and show the results base on EVT analyses for both learning contexts. SE measures the spatial distribution of the gaze of each peer. To compute SE, we first define a 50-pixel by-50-pixel grid over the screen and we compute for each peer the proportion of fixation time located in each grid cell (Figure 3). This proportion is computed over a time window of five seconds. This results in a proportionality matrix and the SE is computed as the Shannon entropy of this 2-dimensional vector. The spatial entropy is also task-independent, as it can be computed for any task, but the interpretation of the entropy values might be dependent on the visual stimuli. A low value of SE would mean that the subject is concentrating on a few elements on the screen, while a high SE value would depict a wider focus size.

What does extreme spatial entropy mean? The idea is to capture the visual focus size (not attention, although in the contexts of the two examples they might be related) of the participants. The higher the spatial entropy is, the larger the focus size is. A spatial entropy value of zero indicates that the participant is looking at only one part of the screen and higher values indicate that the participant looks at different parts of the screen, during a given time window. Now, an extreme spatial entropy would indicate that the participant is looking “*all over the place*”.

## Results

### Collaborative concept map

First, using the univariate EVT, we check the average return levels of the two spatial entropy time series. The average return levels (calculated at 95<sup>th</sup> percentile) for spatial entropy is lower for the pairs with high collaboration outcome ( $F[1,15.78] = 6.53$ ,  $p\text{-value} = .01$ , one-way ANOVA without assuming equal variances) than the average return levels of spatial entropy for the pairs with low collaboration outcome.

Second, considering bivariate EVT results for the spatial entropy of the peers, there are two values to be checked: 1) the level of extremal dependence and 2) the strength of extremal dependence if the level is non-zero. We observe a higher extremal dependence (calculated at the 95% quantile) between the spatial entropy of peers with low collaboration outcome ( $F[1,22] = 4.28$ ,  $p\text{-value} = 0.01$ ) than the extremal dependence between the spatial entropy of peers with high collaboration quality. We observe an even more significant difference in the strength of extremal dependence (calculated at the 95% quantile) for the pairs with the two different levels of collaboration outcome ( $F[1,22] = 10.43$ ,  $p\text{-value} = 0.001$ ). Pairs with low level of collaboration outcome have stronger extremal dependence between the spatial entropy of peers than that for the pairs with high level of collaboration outcome.

### Collaborative intelligent tutoring systems

In the univariate EVT case for the ITS data, the average return levels (calculated at 95<sup>th</sup> percentile) for spatial entropy is negatively correlated with the pair's average posttest score ( $\text{cor} = -0.48$ ,  $p = .01$ ). Next, we look at the bivariate EVT for the ITS data. There is a negative correlation between the upper tail dependence (calculated at the 95% quantile) for the spatial entropy of peers and the average post test score of the pairs ( $\text{cor} = -0.48$ ,  $p = .01$ ). In the case of ITS data, we observe a significant and negative correlation between the strength of upper tail

dependence (calculated at the 95% quantile) for the spatial entropy or pairs and their average posttest score ( $cor = -0.43, p = .03$ ).

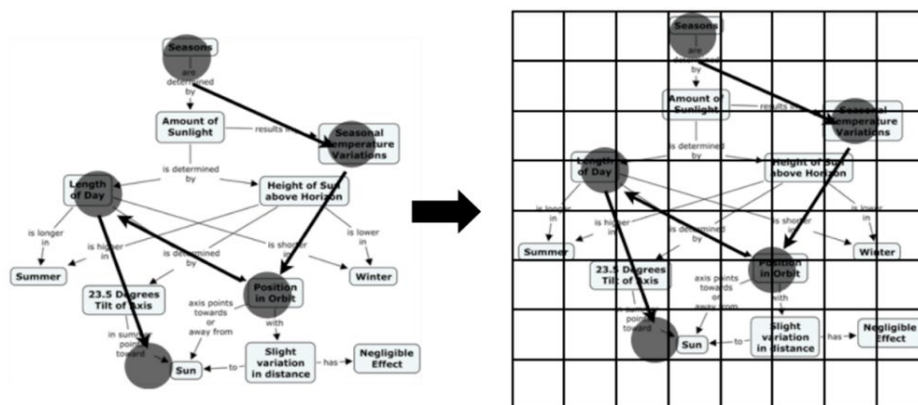


Figure 3. The process of computing entropy. The image on the left shows the exemplar concept-map and gaze patterns (grey circles and arrows). The image on the right shows the placement of the grid.

## Discussion

In this paper, we presented a new method in the context of analysing CSCL data, specifically dual eye-tracking data. We propose that EVT based methods are robust enough to estimate the probabilities of the rare events, which can then be used to provide proactive feedback to students. As examples, we presented results from two dual eye-tracking studies: collaborative concept map and collaborative ITS. To explain the findings from both the studies in a unified way, we use the same metric to capture the focus (spatial entropy) of the peers in two studies.

In the univariate EVT case, we propose to use the return value at the 95th percentile. This value of the spatial entropy has 5% chances of being exceeded. Across both the concept map and ITS contexts, the results indicate that the average values of extreme entropy is higher for the pairs with the low collaborative outcome/learning than the pairs with high collaborative outcome/learning. This simply translates to the fact that pairs with high levels of collaborative outcome/learning have lower levels of spread-out gaze patterns.

The bivariate context provides a supporting explanation for the results from the univariate EVT. We observe that both the level and strength for the upper tail distribution of the spatial entropy is negatively correlated with the collaborative performance/learning, i.e., the moments of extreme entropy appear together in time for the pairs with low collaborative outcome/learning. Combining this with the univariate results, we can conclude that the peers with low collaborative outcome/learning not only have higher chances of “looking all over the place” but there are high chances of them looking all over the place at the same time.

This indicates that the pairs with low collaborative outcome have moments where both the participants in the pair have extremely large visual focus. The cause of such behavior could be explained in two different ways. First, both the participants are looking for some information on the screen and thus they have a large visual focus size. Second, the mutual understanding has a missing link that needs to be created between the two peers. This information can be used to intervene proactively by providing greater scaffolding for student interactions to help guide the collaborative learning process when students have low focus at the same time.

In terms of contemporary dual eye-tracking analyses methods, the bivariate EVT is like analysing the gaze cross-recurrence. Gaze cross recurrence (CR) has been found to be correlated with the collaboration quality (Jermann and Nuessli, 2012; Schneider et.al, 2013). CR indicates the time peers spent looking at the same area on the screen at the same time. In the bivariate EVT case, visual focus of the peers is compared over time. Having comparable visual focus size does not necessarily mean that the peers are looking at the same part of the screen. However, having a large focus size in each time window will result in a high CR over time, as in an aggregated time frame the peers would be looking at the same area on the screen, which is the whole screen. One interesting finding from the examples presented in this paper is the relation between the extreme visual focus and the average learning gains of the peers. In a recent contribution (Olsen, Alevén, & Rummel, 2017), the researchers did not find a relation between CR and overall learning gains of the students using the same collaborative ITS data. This suggests that by analysing the HILF data, we may be able to gain additional insights that are not apparent when analysing the entire data set.

The relation between the extreme visual focus and collaborative outcome/learning, provides an opportunity for designing proactive feedback tools to support collaboration. Using the methods described in this

contribution, one can identify the key moments to provide the feedback to the collaborators. Most of the recent work done in the direction of using gaze awareness to scaffold collaboration has been focused onto showing the gaze of the peers to each other. For example, Ishii and Kobayashi (1992) and Monk and Gale (2002) designed systems displaying the face of the collaborators. Stan and Brennan (2004) showed that displaying partners' gaze while debugging a program helped finding bugs. In another experiment, Brennan, et. al. (2008) showed that displaying partners' gaze in "Os-in-Q" search helped the collaborators in more effective labor division. Recently, the gaze-awareness has been used to support collaboration in high level tasks such as pair programming (D'Angelo and Begel, 2017). Gaze awareness has also been shown to be useful in co-located collaboration scenarios (Van Rheden et. al., 2017) and remote collaborations using different types of devices (Akkil, et, al., 2018). One common theme across these gaze visualisations is that the gaze is visualized throughout the whole collaborative work, which might hinder the learning experience on a few occasions. Identification of the key moments during the collaboration for providing support might improve the effectiveness of the gaze aware tools.

With the results from the two DUET studies based on the EVT, we can estimate a high value of visual focus for both the participants which has a very low probability of being exceeded. In a collaborative scenario if we observe that the visual focus sizes for both the participants in the dyad is going to exceed a certain threshold, we can proactively support the pair to avoid disruptions in the collaboration. EVT has these added values for gaze-aware feedback systems, which could enable the instructor to be proactive and select the exact moments to provide guidance rather than visualizing the support throughout the collaboration.

## Conclusions

We presented a complementary method to analyse CSCL data based on the tails (lower or upper) of the distribution. This method is based on a mathematical theory, which is novel to CSCL community, called Extreme value theory. The main motivation behind using EVT is to be able to provide proactive scaffolding during the moments, when the collaboration between peers is at a point where the outcome/quality is disruptive. We do not claim the superiority of this method (for a comparison with the traditional methods of analyses, see Sharma et. al., 2016). We propose, that EVT provides a different point of view for the data when the traditional methods do not apply because of the failed assumptions or when the traditional methods do not provide any useful insights about the collaborative processes, outcome or quality.

Within this paper, we provide a contribution both through the application of EVT to educational process data as well as furthering the understanding of how dual eye-tracking relates to collaborative learning outcome measures. By analysing just, the tail of the data, we can distinguish patterns that may not have otherwise been apparent. Across contexts, we have shown a common pattern of with low outcome measures having lower focus at the same time. We propose EVT as an alternative method for the analysis of collaborative learning data that can provide complementary viewpoints to the common CSCL methodological repertoire.

## References

- Akkil, D., & Isokoski, P. (2018). I see what you see: gaze awareness in mobile video collaboration. In Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications (p. 32). ACM.
- Allopenna, P.D., Magnuson, J.S, and Tanenhaus, M.K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models\* 1,\* 2,\* 3,\* 4,\* 5. *Journal of memory and language*, 38(4),
- Bednarik, R., & Kauppinen, M. (2013). Unravelling the interaction strategies and gaze in collaborative learning with online video lectures. In Proceedings of the 6th workshop on Eye gaze in intelligent human machine interaction: gaze in multimodal interaction (pp. 57-62). ACM.
- Belenky, D., Ringenberg, M., Olsen, J., Alevan, V., & Rummel, N. (2014). Using Dual Eye-Tracking to Evaluate Students' Collaboration with an Intelligent Tutoring System for Elementary-Level Fractions. Grantee Submission.
- Coles, S. (2001). *An introduction to statistical modeling of extreme values*. London: Springer. <http://dx.doi.org/10.1007/978-1-4471-3675-0>
- D'Angelo, S., & Begel, A. (2017). Improving communication between pair programmers using shared gaze awareness. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (pp. 6245-6290). ACM.
- Deiglmayr, A., Rummel, N., & Loibl, K. (2015). The Mediating Role of Interactive Learning Activities in CSCL: An Input-Process-Outcome Model In Lindwall, O., Häkkinen, P., Koschman, T. Tchounikine, P. Ludvigsen, S. (Eds.) (2015). *Exploring the Material Conditions of Learning: The Computer Supported Collaborative Learning (CSCL) Conference 2015, Volume 2*. Gothenburg, Sweden: The International Society of the Learning Sciences.

- Embrechts, P. (2000). Extreme Value Theory: Potentials and Limitations as an Integrated Risk Management Tool. Manuscript, Zurich, Switzerland: Department of Mathematics, ETH, Swiss Federal Technical University.
- Fussell, S. R., Setlock, L. D., & Parker, E. M. (2003). Where do helpers look?: gaze targets during collaborative physical tasks. In CHI'03 Extended Abstracts on Human Factors in Computing Systems (pp. 768-769). ACM.
- Griffin, Z.M., and Bock, K. (2000). What the eyes say about speaking. *Psychological science*, 11(4).
- Ishii, H. and Kobayashi, M. (1992). Clearboard: a seamless medium for shared drawing and conversation with eye contact. In Proceedings of CHI, ACM.
- Jermann, P., & Nüssli, M. A. (2012). Effects of sharing text selections on gaze cross-recurrence and interaction quality in a pair programming task. In Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work (pp. 1125-1134). ACM.
- Monk, A.F., and Gale, C. (2002). A look is worth a thousand words: full gaze awareness in video-mediated conversation. *Discourse Processes*, 33(3), 257-278.
- Olsen, J. K., Alevan, V., & Rummel, N. (2017). Exploring Dual Eye Tracking as a Tool to Assess Collaboration. In *Innovative Assessment of Collaboration* (pp. 157-172). Springer, Cham.
- Olsen, J., Sharma, K., Alevan, V., & Rummel, N. (2018). Combining Gaze, Dialogue, and Action from a Collaborative Intelligent Tutoring System to Inform Student Learning Processes. International Society of the Learning Sciences, Inc.[ISLS].
- Papavlasopoulou, S., Sharma, K., Giannakos, M., & Jaccheri, L. (2017). Using Eye-Tracking to Unveil Differences Between Kids and Teens in Coding Activities. In Proceedings of the 2017 Conference on Interaction Design and Children(pp. 171-181). ACM.
- Prieto, L. P., Sharma, K., & Dillenbourg, P. (2015). Studying teacher orchestration load in technology-enhanced classrooms. In *Design for teaching and learning in a networked world* (pp. 268-281). Springer, Cham.
- Ramesh, N. I., & Davison, A. C. (2002). Local models for exploratory analysis of hydrological extremes. *Journal of Hydrology*, 256(1-2), 106-119.
- Richardson, D. C., Dale, R., and Kirkham, N. Z., (2007). The art of conversation is coordination. *Psychological Science*, 18(5).
- Raca, M., & Dillenbourg, P. (2013, April). System for assessing classroom attention. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 265-269). ACM.
- Santinelli, L., Morio, J., Dufour, G., & Jacquemart, D. (2014). On the sustainability of the extreme value theory for WCET estimation. In *OASIS-OpenAccess Series in Informatics* (Vol. 39). Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.
- Schneider, B., Abu-El-Haija, S., Reesman, J., & Pea, R. (2013). Toward collaboration sensing: applying network analysis techniques to collaborative eye-tracking data. In Proceedings of the Third International Conference on Learning Analytics and Knowledge (pp. 107-111). ACM.
- Schneider, B., Sharma, K., Cuendet, S., Zufferey, G., Dillenbourg, P., & Pea, R. (2016). Using mobile eye-trackers to unpack the perceptual benefits of a tangible user interface for collaborative learning. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 23(6), 39.
- Sharma, K., Chavez-Demoulin, V., & Dillenbourg, P. (2016). An Application of Extreme Value Theory to Learning Analytics: Predicting Collaboration Outcome from Eye-tracking Data.
- Sharma, K., Alavi, H. S., Jermann, P., & Dillenbourg, P. (2016). A gaze-based learning analytics model: in-video visual feedback to improve learner's attention in MOOCs. In Proceedings of the Sixth International Conference on Learning Analytics & Knowledge (pp. 417-421). ACM.
- Sharma, K., Jermann, P., Nüssli, M. A., & Dillenbourg, P. (2013). Understanding collaborative program comprehension: Interlacing gaze and dialogues. In Proceedings of Computer Supported Collaborative Learning (CSCL 2013) (Vol. 1, No. EPFL-CONF-184007, pp. 430-437).
- Sharma, K., Caballero, D., Verma, H., Jermann, P., & Dillenbourg, P. (2015). Looking AT versus looking THROUGH: A dual eye-tracking study in MOOC context. International Society of the Learning Sciences, Inc.[ISLS].
- Sharma, K., Olsen, J. K., Alevan, V., & Rummel, N. (2018). Exploring Causality Within Collaborative Problem Solving Using Eye-Tracking. In *European Conference on Technology Enhanced Learning* (pp. 412-426). Springer, Cham.
- Smith, R. L. (1990). Extreme value theory. *Handbook of applicable mathematics*, 7, 437-471.
- Van Rheden, V., Maurer, B., Smit, D., Murer, M., & Tscheligi, M. (2017). LaserViz: Shared gaze in the Co-located physical world. In Proceedings of the Eleventh International Conference on Tangible, Embedded, and Embodied Interaction (pp. 191-196). ACM.

# Exploring Disciplinary Boundaries in Early Elementary Students' Developing Practices

Kylie Pepler, University of California Irvine, kpepler@uci.edu

Joshua Danish, Indiana University, jdanish@indiana.edu

Naomi Thompson, Indiana University, naothomp@indiana.edu

**Abstract:** In CSCL we often refer to knowledge in terms of “practices” to highlight how knowledge is performed in context rather than abstracted from it (e.g., Greeno, 1998). One aspect of context is the notion of disciplines such as science, math, and the language arts (Sawyer, 2006). To date, however, research that explores the relationship between knowledge and disciplinary context has typically focused within a single disciplinary frame, such as science (e.g., Banks et al., 2007). This paper aims to explore the influence of disciplinary boundaries between science and language arts on students' developing practices. We present data from a mixed-age first and second grade classroom where students collaboratively engaged with a book about honeybees from either a science or a language arts frame. Analyses of students' individual written critiques reveal how they consistently adapted their criteria for what makes a good representation based upon the disciplinary framing.

## Introduction

Core to CSCL and the Learning Sciences is the belief that context matters (Sawyer, 2006) and that the social, cultural, and physical environment in which students learn new content fundamentally shapes what they have learned (Greeno, 2006). For this reason, sociocultural theorists often refer to knowledge in terms of the “practices” that students know how to engage in, to highlight the situated and interactive nature of knowledge, and how knowledge is performed in context rather than abstracted from context (Greeno, 1998, 2011; Lobato, 2012).

In an effort to understand the relationship between learning and context, researchers have focused on many aspects of learning environments including whether students are working individually or in groups (e.g., Webb et al., 2009) and whether students are learning in or out of school (Kafai, Pepler, & Chapman, 2009) to name a few. However, one aspect of context, which cuts across all of these other variables and remains crucial for supporting learning, is the notion of disciplines such as science, math, and the language arts (Sawyer, 2006). To date, however, research that explores the relationship between knowledge and disciplinary context has typically focused within a single disciplinary frame, such as science, even when it explores students' developing practices within disparate social contexts such as the school and home (Banks et al., 2007). Furthermore, while there is an abundance of work examining STEM (Science, Technology, Engineering, and Math) disciplines, other disciplinary contexts such as the language arts are largely overlooked. The question then emerges, how do these disciplinary contexts influence students' practices?

Therefore, this paper aims to explore the influence of disciplinary boundaries between science and language arts on students' developing practices. We selected students early in their academic career, in first and second grade, to highlight and test whether students' awareness of disciplinary boundaries develops early in the process of schooling. Students in elementary school classrooms typically remain in the same room and work with the same teacher as they transition between disciplines, contrasted with middle and high school students who frequently move to a new classroom and work with a different teacher for each discipline. One might then assume that disciplinary boundaries are less salient at this stage of schooling, although our experiences suggest that this is not the case. Further, educators typically assume that students can learn discipline-specific practices in a variety of disciplinary contexts, and that this is likely to promote more robust understanding of the links between disciplines. For example, many of our collaborating teachers have asked students to read about the current science topic during their literacy time, assuming that the students will learn the science content and the reading practices simultaneously. However, researchers have rarely studied whether this assumption is true. By exploring disciplinary boundaries at this stage of schooling, and within a context where disciplinary boundaries might be more subtle, we aim to demonstrate how salient these boundaries actually are for students, and how they impact student learning and knowing. Our core research question is: *To what extent does explicit disciplinary framing of an activity as science or language arts affect young children's classroom practices?*

We present data from a mixed-age first and second grade classroom where students were asked to evaluate a short book about honeybees before and after extensive language arts and science instruction. This activity was chosen because it required students to make their disciplinary values explicit, allowing us to contrast the influence of the two disciplinary contexts and the practices that students engage in within those contexts.

Students were randomly assigned to one of two conditions where the honeybee book was framed as either a “science” representation or as a “language arts” representation. Students then completed several activities to give them a range of opportunities to reveal their perspective on what makes a “good” or accurate representation within the specific disciplinary context. The students in each group then completed an identical 10-week science curriculum where they learned about how honeybees collect nectar (Danish, Pepler, Phelps, & Washington, 2011) as well as standard language arts instruction on grammar, spelling, and strategies for good writing/storytelling before again critiquing storybooks about bees. Close analyses of students’ written critiques reveal how they consistently adapted their criteria for what makes a good representation based upon the disciplinary framing, often overlooking errors in one disciplinary condition that were treated as crucial in the other.

### **Contrasting disciplinary practices using student representations**

Our work is grounded in sociocultural theories of learning (Cole, 1996; Greeno, 2006, 2011). These theoretical frameworks share a common ancestry in Vygotsky (1978) and highlight the role of the social environment in learning and development. From this perspective, disciplinary knowledge includes not only specific facts but practices of engaging with content, people, and materials (Greeno, 2011), and even disciplinary ways of perceiving the world around us (Goodwin, 1994; Stevens & Hall, 1998). As such, the notion of “discipline” has emerged as crucial to structuring how we think about learning environments, and how we evaluate students’ knowledge (Sawyer, 2006).

Thus, our goal in the current study is to examine students’ participation within disciplinary activities in order to begin identifying the relationship between their shared and individual understandings of what the disciplinary context means, while also teasing out the role of the environment in shaping those perspectives. To ground our contrast we identified a common practice of both science and language arts—the critique, or evaluation of quality, of visual representations.

### **Representations as a site for disciplinary comparison**

Representations, such as drawings and narrative storylines, play a role across disciplines and contexts (diSessa, 2004; Lehrer & Schauble, 2000; Schwartz & Heiser, 2006) and yet serve a different purpose within each discipline. For example, a drawing of a flower might be used as an anatomical reference in a science classroom, and need to be accurate. In contrast, a drawing of a flower in language arts might instead be a starting point for a new story. Representations are also commonplace in early elementary classrooms because they make it easy to capture, share, and relate information in powerful ways (Schwartz & Heiser, 2006; Willats, 2005). Students are frequently asked to create or work with representations as part of their daily activities.

Much of the research into students’ engagement with representations, including their critiques, has focused on either the cognitive benefits or abilities of individual students (Schwartz & Heiser, 2006), or a specific disciplinary context such as science or math (Lehrer & Schauble, 2005). However, an increasing body of research has drawn upon sociocultural theories of learning to move beyond individual cognition and explore the relationship between representations and the activities in which they are created, modified, and used (Hall, 1996; Roth, 1997; Roth & McGinn, 1998). These sociocultural studies accomplish this by examining observable patterns in student behavior (practices) as they develop over time, or the immediate influence of a specific activity upon students’ actions. We blend these two approaches by comparing students’ practices not only across two distinct contexts, but also as they change over time. Furthermore, despite an interest in context, prior sociocultural studies continue to focus largely on representations in math and science. Our present study aims to extend this pattern by contrasting students’ creation of representations in a science context with a language arts context. Language arts was chosen because we hypothesized that it would be a distinct enough setting from science to make students’ awareness of disciplinary contexts visible.

### **Representational critique in science and language arts**

Representations play an important and unique role in science as a discipline, allowing researchers to share theories, experiments, and evidence across vast distances and different time periods (Latour, 1987, 1988; Roth & McGinn, 1998). When working with representations, one important practice shared by multiple disciplines is the ability to critique and evaluate representations, both those that others have produced as well as one’s own, in an effort to improve and refine it (diSessa, 2002). This practice of critique is also instrumental in providing students with an opportunity to develop and explore their own identity with respect to the representation and the discipline (Greeno, 2011). Effective critique of a representation requires an understanding of the purpose of the representation and the context of its use. Fortunately, research shows that young children (i.e., kindergarten through second grade) are capable of critiquing representations along a number of dimensions, including the



accuracy of their content (Danish & Enyedy, 2007; Danish & Phelps, 2010a). Critique can be particularly powerful as students create their own representations, allowing them to explore both the representational form and the science content referred to by the representation (Danish & Phelps, 2010b; Parnafes, 2010).

While there is significantly less existing research investigating the nature of representations within the context of language arts, we do know that language arts representations, especially storybooks, videos, and other media, play a critical role in the development of language arts skills. We looked to the Common Core standards for our initial criteria. To further identify language arts practices which are likely present when engaging with science content, we also focus on informational and technical writing skills which are relevant to all STEM fields and critical 21st century skills (Duke, 2010; Purcell-Gates, Duke, & Martineau, 2007). Building on the idea of informational writing, we then generated a grounded list of criteria that resulted in a coding scheme that emphasized technical writing, grammar (such as correct capitalization, verb form, etc.) and the construction of a clear story-arc (with a beginning, middle and end to the story), among other criteria further explored in the methods section below.

## Methodology

### Participants

The present study took place in a public Midwestern elementary school with 20 first and 20 second grade students (ages 6-9; 40 total) in a mixed-age classroom, with 37 of the students being present at the two time points in which data was collected and included in the present study. The majority of children were White (90%) with 17 % of the students receiving free or reduced lunch. Students were randomly divided into two conditions: science (N = 19) and language arts (N = 18).

### Procedure

The activities in the two conditions (and at both time points) were identical except that the activities in one were framed by the teacher as “science” and in the other as “language arts.” The mixed-age classroom is co-taught by two teachers. One of these teachers led the science condition, and the other led the language arts condition. Steps were taken to ensure that as much as possible, the sequence and the methods of the two conditions were the same by scripting the directions and questions in advance of the conversation. Within each condition, the teacher followed a predetermined sequence for the 50 minute session: 1) generate definitions for what makes a good representation in the assigned disciplinary context (i.e., science or language arts); 2) record the groups’ thoughts on good representations on a whiteboard; 3) introduce a short storybook about a honeybee finding and collecting nectar; 4) ask students to write or draw on a copy of the book about what made it a good or bad representation 5) follow-up whole group discussion of individual observations and thoughts about what made the book either good or bad; 6) discuss a revised version of the book with many intentional errors removed to see if students felt it was “better” or not.

The research team created two storybooks for these activities (one for use at each time-point) representing a honeybee collecting nectar, one for use at each time point. In both cases, we intentionally violated the same number of scientific principles (e.g., the number of anatomically incorrect bees) and language arts principles (e.g., counts of missing or inaccurate punctuation) in an effort to provide approximately equivalent opportunities for students to engage in the practice of discipline-specific critique. We provided hand-drawn images and hand-written text for the students to feel more comfortable in the critique process (to appear more like a peer seeking feedback on an early draft rather than a published text). Data were collected as part of the larger BeeSign study which indicated that students in both conditions learned extensive and high-level science content over the course of the study (Danish, Pepler, Phelps, & Washington, 2011; Pepler & Danish, 2012). The knowledge and understanding built during the classroom activities were largely group-based and collaborative. Here, we analyze individual measures to better understand the impact of the group and the environment on individual learning.

### Analysis

To summarize and contrast students’ critical practices in the different conditions, student utterances in whole-group discussion and their writings (including the drawings, text, or other marks on the photocopied storybook) were collected, transcribed, and coded to reflect the content of their critique as relating to science (e.g., accuracy, parsimony), language arts (e.g., grammar, storyline), or being non-domain-specific (e.g., aesthetic preference for color over black and white drawings). This paper reflects findings from students’ written critiques. Additionally, an average percent of time spent engaging science, language arts, and other was calculated per child to account for the wide variability in writing ability at early ages. One researcher coded all of students’ written comments,

with a second researcher coding a randomly selected 30% of all data to establish inter-rater reliability, resulting in high agreement (language arts kappa  $\kappa=.691$ , science  $\kappa=.898$ ).

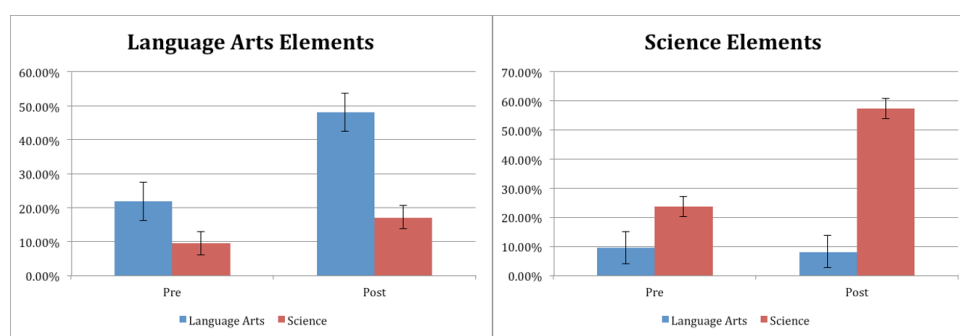
To explore the relationship between time and condition upon the mention of science or language arts content in the written critiques, a split-plot analysis of variance (ANOVA) was performed. Our hypothesis was that regardless of the disciplinary context framing students would pull from their science repertoire to create and talk about their storyboards due to their participation in the extensive science unit that targeted the same content represented in the honeybee books. A chi-square analysis was also completed to determine whether there was a significant difference in the proportion of each topic being mentioned across conditions and time-points.

## Results

Across both conditions, students generally provided more discipline-specific critiques in the post-condition than the pre. This is unsurprising and likely attributable to the time spent learning the respective disciplinary practices in the intervening weeks. Below we describe the shifts in written critique practices across conditions.

### Written critiques

Across both conditions, students' written critiques reflected an impact of both time point and disciplinary framing. Students were less likely than hypothesized to bring their practices of science critique into the language arts condition and vice-versa (see figure 1 and table 1). These results are discussed in further detail below.



**Figure 1.** Percentage of written representational critique elements coded as language arts (left) and science (right). Remaining percentage refers to items coded as non-specialized or other.

**Table 1:** Mean number of written critique elements coded as language arts and science

|               |    | Language Arts Codes |                 |              |                  | Science Codes |                  |              |                  |
|---------------|----|---------------------|-----------------|--------------|------------------|---------------|------------------|--------------|------------------|
|               |    | Pre                 |                 | Post         |                  | Pre           |                  | Post         |                  |
| Condition     | N  | M<br>(count)        | %<br>(SD)       | M<br>(count) | %<br>(SD)        | M<br>(count)  | %<br>(SD)        | M<br>(count) | %<br>(SD)        |
| Language Arts | 18 | 21.89%              | 16.82<br>(3.83) | 47.98%       | 23.45%<br>(6.06) | 9.41%         | 10.55%<br>(1.05) | 8.08%        | 10.08%<br>(0.79) |
| Science       | 19 | 9.49%               | 8.76%<br>(1.83) | 17.09%       | 14.66%<br>(2.72) | 23.79%        | 18.75%<br>(3.05) | 57.36%       | 19.81%<br>(6.11) |

### Language arts elements

Within the language arts group, there was an overall increase in the use of language arts elements in students' written critiques, from about 22% to about 48% of the students' written critiques. In the science group, there was also a (much smaller) increase in written critique of language arts elements, from about 9% to about 17%. (see both figure 1 and table 1). Although both groups saw their scores increase, the increase in the language arts group was proportionally more than three times the increase in the science group. This was a statistically significant interaction between time of measurement and context framing,  $F(1, 35) = 21.03$ ,  $p < .001$ . The strength of this relationship, as assessed by partial  $\eta^2$ , was strong, with the interaction factor accounting for 37.5% of the variance in language arts elements.

Results also suggest that there was a main effect for both condition and time point such that the frequency of language arts elements including both pre and post was higher in the language arts condition compared to the science condition (32% vs. 13%) and that more language arts elements were present in the post-test than in the pre-test across both conditions (30% vs. 15%).

Taken together, these results suggest two key trends in students' practices for written language arts critique. First, the disciplinary framing of the activity had a dramatic impact upon students' practices, leading students to engage far more consistently on the language arts critiques during the language arts condition. Second, this effect appears to have been magnified over time. Thus the students in the science condition provided only marginally more critiques during the post-activity whereas the students in the language arts condition increased dramatically in the number of written language arts critiques they included. This increase can be attributed to continued engagement in the language arts curriculum during the intervention. Because all of the students experienced an identical classroom language arts curriculum over the 10-week period, we do not believe that these differences were due to chance but rather to the disciplinary framing of the activity. This highlights a key interaction between students' developing practices and disciplinary framing given that students' shift in practices related to language arts were much more evident in the language arts context.

As we turn to what this looked like for individual students in each of these conditions, Figure 2 illustrates the increase in language arts elements from the pre to post intervention written critiques. These examples are representative of the kinds of critiques the students generally provided and highlight the marked changes that occurred between the pre and post intervention in the language arts condition.

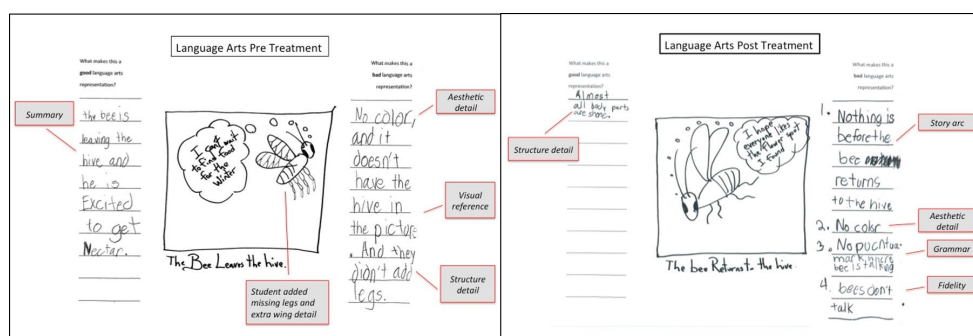


Figure 2. Sample language arts pre-intervention critique from Student 1 (left), sample language arts post-intervention critique from Student 1 (right).

In Figure 2 (left), the participant in the Language Art pre-intervention focused on general features of the image. When prompted, “what makes this a good language arts representation?” the student gave a summary response about the bee leaving the hive and being excited to get nectar, stating that “the bee is leaving the hive and he is Excited [sic] to get nectar”. This was coded as “other.” When prompted, “what makes this a bad language arts representation?” the student focused largely on the picture rather than the words, and used both the written and visual edits to identify some of the intentional errors in the structural details of the drawing of the bee. In this example, Student 1 drew on the booklet, adding the missing legs to the bee and adding details to the wings which was coded as “science” as it added greater scientific fidelity to the bee. Notably, the student has ignored other intentional grammar violations on the page, such as the uppercase ‘B’ in bee and ‘L’ in leaves as well as other intentional science and language arts violations that were included in the storyboard, including other inaccurate body parts (such as the missing thorax of the bee).

When comparing Student 1’s critique in the post-test condition after the 10-week curriculum (see Figure 2, right), several changes are apparent. There are still some general or science critiques found in the storyboard, such as reference to the fact that, “almost all of the body parts are [shown],” (coded as structure detail important to science), For example, the student critiques the story arc of the drawing because there is, “nothing before the bee returns to the hive.” The student also makes clear reference to the punctuation on the page, noting that there is, “no punctuation mark where bee is talking.” The student did not, however, pick up on the intentional violation of grammar in the text on the page with an upper case ‘R’ in the word returns.

### Science elements

There was also a significant interaction between time of measurement and context framing in the prevalence of students’ written science critiques,  $F(1, 35) = 13.60, p < .001$ . The strength of the relationship, as assessed by partial  $\eta^2$ , was strong, with the interaction factor accounting for 27.9% of the variance in written science elements. There was a slight decrease in the mean percentage of science elements that each student included in their written critiques, from 9% to 8%, in the language arts context, and a significant increase in science elements in the science context, from an initial 23% to 57% of the written student critiques focusing on the critique of science elements (see table 1).

The frequency of science elements was also higher in the science context compared to the language arts condition (40% vs. 8% of the writings) and that more science elements were present in the post-test than in the pre-test (32% vs. 16%). This suggests that the context framing again had an impact in the type of criteria children picked to critique the storyboards, with students increasingly referring to science criteria as they learn them, particularly in the science context. Thus, the findings for students' practices of critiquing the science content of the representations appear to generally parallel those for language arts, reflecting an impact of both context and learning over time.

The case shown in figure 3 illustrates the findings from the science group before and after the 10-week curriculum. As with the language arts critique, this example was representative of the critiques of the classroom as a whole and particularly highlights the marked changes in the critique from the science condition from pre to post intervention.

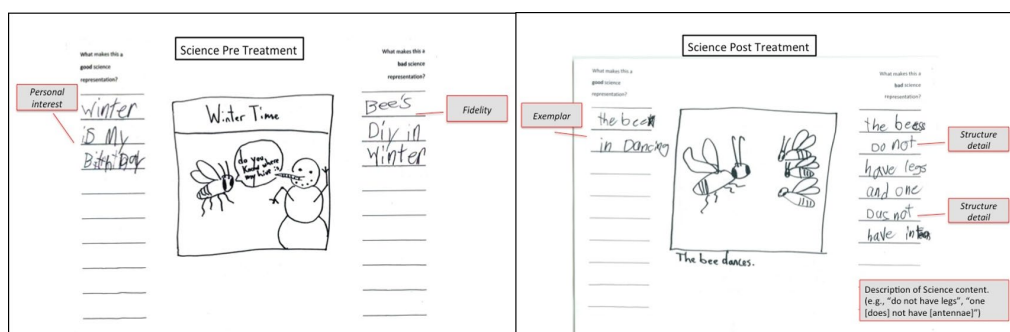


Figure 3. Sample science pre-intervention critique for Student 2 (left), sample science post-intervention critique from Student 2 (right).

In figure 3 (left), when the Student 2 was asked, “what makes this a good science representation?” the student response was simply a statement of personal interest (“Winter is my birthday”) and thus was coded for “other.” In response to, “what makes this a bad science representation?” the student makes reference to the fact that, “bees [die] in winter.” While this is not an entirely accurate understanding of how bees survive in the winter (as many survive within the hive), it demonstrated the value that the student placed on science fidelity that was an example of a science code.

In the post-intervention critique from the same Student 2, the student noted that this is good because it is a representation of the ‘waggle dance.’ This was coded as a science exemplar as the waggle dance was a central theme of the curriculum of how bees communicate. The student also showed marked improvement on recognizing missing science elements such as, “the bees do not have legs and one [does] not have [antennae]”, which were both coded as structure details important to science (see figure 3, right).

## Discussion

Our analyses suggest that students as young as first grade are well attuned to the particular disciplinary frame in which they work as they critique representations. In fact, they are quite likely to ignore major flaws in sample representations when evaluating it in a content domain in which those flaws are not central (e.g., students ignored grammatical errors when evaluating a representation framed as “scientific”). In short, our results indicate that disciplinary context plays a major role in shaping students’ engagement with representational products, and that students’ engagement in disciplinary content may at times be seen across the curriculum. While this paper focused on the students’ written critiques, separate analyses of the verbal discussions suggested this trend continues when the whole group is in conversation. Given the increasing interest in cross-curricular activities designed to help students make connections throughout their learning activities, this suggests important opportunities for reflecting upon the quality and consistency of students’ representational materials and activities. As the current piece does not dive into the interactional processes of learning, these findings also suggest the necessity of future research to more explicitly explore the impact of disciplinary framing upon student learning, particularly in these cross-curricular contexts. These collaborative micro-processes will continue to be of interest and importance to the CSCL community.

While these findings generally support the situative or sociocultural view that context matters in exploring cognition and learning, the details of our study offer several additional suggestions for researchers and practitioners. First, they point to some very real challenges inherent in cross-curricular activity designs. Increasingly, there is a push in early elementary education to enact interdisciplinary activities designed to help

students see the connections across domains. In our work, we have found, for example, that as soon as the teachers know we are interested in teaching bees during “science time”, they identify books about bees for reading time, and have even worked with the art teacher to have students draw bees in art class. And yet, our findings suggest that these bridges are even harder to build than most educators likely anticipate. Given that the students in our experiment focused so intently upon language arts criteria as they critiqued science books within the language-arts-framed condition, it raises a question regarding what students might learn when reading about science topics in a time that is clearly designated as language arts? An important next-step in this line of research will be to examine how disciplinary framing of this nature influences student learning and other kinds of practices (e.g., creating representations rather than just critiquing them) to help address this question and effective practices to help bridge these disciplinary boundaries that seem to be so salient at even young ages.

A second and related result of this work is that there is clearly value in attending much more closely to how teachers frame activities as linked to a specific discipline, and what impact this has upon students. This is particularly relevant in early elementary classrooms where students typically work with the same teacher across content areas, as opposed to moving to new rooms and teachers as they frequently do in secondary school, and thus the shift between disciplines is marked most clearly by the teacher. In our experiment the teacher clearly framed the storybook activity as science or language arts, and then maintained this frame in follow-up questions. What is less clear is which features of the teachers’ prompts the students found salient in determining which practices to perform within the context. Future work which teases this out will be an important first step in understanding how to help students recognize the disciplinary context in which they are operation in order to promote discipline-specific practices.

## References

- Banks, J. A., Au, K. H., Ball, A. F., Bell, P., Gordon, E. W., Gutiérrez, K. D., . . . Mahiri, J. (2007). Learning in and out of school in diverse environments: Life-long, life-wide, life-deep: LIFE Center, University of Washington, Stanford University, and SRI International.
- Cole, M. (1996). Cultural psychology: a once and future discipline. Cambridge, Mass.: Belknap Press of Harvard University Press.
- Danish, J. A., & Enyedy, N. (2007). Negotiated Representational Mediators: How Young Children Decide What to Include in Their Science Representations. *Science Education*, 91(1), 1-35.
- Danish, J. A., Pepler, K., Phelps, D., & Washington, D. (2011). Life in the Hive: Supporting Inquiry into Complexity Within the Zone of Proximal Development. *Journal of Science Education and Technology*, 20(5), 454-467. doi: 10.1007/s10956-011-9313-4.
- Danish, J. A., & Phelps, D. (2010a). Kindergarten and First-Grade Students’ Representational Practices While Creating Storyboards of Honeybees Collecting Nectar. In K. Gomez, L. Lyons & J. Radinsky (Eds.), *Learning in the Disciplines: Proceedings of the 9th International Conference of the Learning Sciences (ICLS 2010) - Volume 1, Full Papers* (pp. 420-427). Chicago IL: International Society of the Learning Sciences.
- Danish, J. A., & Phelps, D. (2010b). Representational Practices by The Numbers: How Kindergarten and First-Grade Students Create, Evaluate, and Modify Their Science Representations. *International Journal of Science Education*. doi: 10.1080/09500693.2010.525798.
- diSessa, A. A. (2002). Students' criteria for representational adequacy. In K. P. Gravemeijer, R. Lehrer, B. V. Oers & L. Verschaffel (Eds.), *Symbolizing, Modeling and Tool Use in Mathematics Education* (pp. 105-130).
- diSessa, A. A. (2004). Meta-Representation: Native Competence and Targets for Instruction. *Cognition and Instruction*, 22(3), 293-331.
- Duke, K. (2010, July 19-24). Teaching Reading and Writing of Procedural or How-to Text. Paper presented at the US Department of Education Reading Institute, Anaheim, CA. .
- Goodwin, C. (1994). Professional Vision. *American Anthropologist*, 96(3), 606-633.
- Greeno, J. G. (1998). The Situativity of Knowing, Learning, and Research. *American Psychologist*.
- Greeno, J. G. (2006). Learning in Activity. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 79-96). New York: Cambridge Univ Press.
- Greeno, J. G. (2011). A situative perspective on cognition and learning in interaction Theories of learning and studies of instructional practice (pp. 41-71): Springer.
- Hall, R. (1996). Representation as Shared Activity: Situated Cognition and Dewey's Cartography of Experience. *Journal of the Learning Sciences*, 5(3), 209-238.
- Kafai, Y.B., Pepler, K., & Chapman, R. (Eds.) (2009). *The computer clubhouse: Creativity and constructionism in youth communities*. New York, NY: Teachers College Press.

- Latour, B. (1987). *Science in Action: How to Follow Scientists and Engineers through Society*. Cambridge, MA: Harvard University Press.
- Latour, B. (1988). Drawing Things Together. In M. Lynch & S. Woolgar (Eds.), *Representation in Scientific Practice* (pp. 19-68). Cambridge MA: MIT Press.
- Lehrer, R., & Schauble, L. (2000). Developing Model-Based Reasoning in Mathematics and Science. *Journal of Applied Developmental Psychology*, 21(1), 39-48.
- Lehrer, R., & Schauble, L. (2005). Developing Modeling and Argument in the Elementary Grades. In T. A. Romberg, Carpenter, T. P., & Dremock, F. (Ed.), *Understanding mathematics and science matters* (pp. 29-53). New Jersey: Lawrence Erlbaum.
- Lobato, J. (2012). The Actor-Oriented Transfer Perspective and Its Contributions to Educational Research and Practice. *Educational Psychologist*, 47(3), 232-247. doi: 10.1080/00461520.2012.693353
- Parnafes, O. (2010). Representational practices in the activity of student-generated representations (SGR) for promoting conceptual understanding. In K. Gomez, L. Lyons & J. Radinsky (Eds.), *Learning in the Disciplines: Proceedings of the 9th International Conference of the Learning Sciences (ICLS 2010) - Volume 1, Full Papers* (pp. 301-308). Chicago IL: International Society of the Learning Sciences.
- Peppler, K., & Danish, J. A. (2012). *E-Textiles for Educators: Participatory Simulations With E-Puppetry*. Paper presented at the The Annual Meeting of the American Educational Research Association, Vancouver, British Columbia.
- Purcell-Gates, V., Duke, K., & Martineau, J. A. (2007). Learning to read and write genre-specific text: Roles of authentic experience and explicit teaching. *Reading Research Quarterly*(42), 8-45.
- Roth, W.-M. (1997). Graphing: Cognitive Ability or Practice? *Science Education*(81), 91-106.
- Roth, W.-M., & McGinn, M. K. (1998). Inscriptions: Toward a Theory of Representing as Social Practice. *Review of Educational Research*, 68(1), 35-59.
- Sawyer, R. K. (2006). *The Cambridge handbook of the learning sciences*. New York: Cambridge Univ Press.
- Schwartz, D. L., & Heiser, J. (2006). Spatial representations and imagery in learning. *The Cambridge handbook of the learning sciences*, 283-298.
- Stevens, R., & Hall, R. (1998). Disciplined perception: Learning to see in technoscience. In M. Lampert & M. L. Blunk (Eds.), *Talking mathematics in school: Studies of teaching and learning* (pp. 107-149). Cambridge: Cambridge University Press.
- Vygotsky, L. S. (1978). *Mind in society: the development of higher psychological processes*. Cambridge: Harvard University Press.
- Webb, N. M., Franke, M. L., De, T., Chan, A. G., Freund, D., Shein, P., & Melkonian, D. K. (2009). 'Explain to your partner': teachers' instructional practices and students' dialogue in small groups. *Cambridge Journal of Education*, 39(1), 49-70.
- Willats, J. (2005). *Making sense of children's drawings*. Mahwah, N.J.: L. Erlbaum Associates.

## Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 1324047 awarded to Kylie Peppler, Joshua Danish, and Armin Moczek. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Thank you to Janis Watson, David Phelps, Justin Whiting, Alejandro Andrade, the teachers who worked with us on this project, and the students who made this work possible.

# Assessing Collaborative Problem Solving: What and How?

Haley W. C. Tsang, Seung Won Park, Laure L. Chen, and Nancy W. Y. Law  
haleywct@connect.hku.hk, swonpark@hku.hk, cllaure@connect.hku.hk, nlaw@hku.hk  
The University of Hong Kong

**Abstract:** With the increasing interests in and recognition of the importance of CPS competence as an educational outcome goal, there will be stronger demands for valid instruments and metrics for the measurement of Collaborative Problem Solving Skills (CPS) from policy makers to education practitioners and parents. The availability of valid instruments and norms for CPS achievement would also promote CSCL adoption within the wider education arena. This study explores two methods of assessing CPS: using the ATC21S online tasks and using qualitative discourse analysis of the same groups of students' transactions as they work together on an authentic learning task. By comparing the measurements arising from both methods, we raise issues and contradictions observed and suggest further research directions related to the measurement of CPS.

## Introduction

Collaborative problem solving (CPS) has been identified as an important competence critical for life and work in the 21st century in the education literature (Fiore et al., 2017) and in education policy documents (e.g. Singapore Ministry of Education (2009)). This is similar to what is stated in the P21 Framework for 21st Century Learning which is developed with participation from educators, business leaders and other stakeholders (Partnership for 21st Century Skills, 2009).

Collaborative learning (including CSCL) have gained increasing prominence as a field of study in education, and the general consensus on the importance of CPS skills, methods and tools for assessing CPS are limited. There are two major, popularly known CPS assessment frameworks and systems in education: the one developed by the ATC21S Consortium (Griffin & Care, 2015; Hesse, Care, Buder, Sassenberg, & Griffin, 2015), and the one used in the assessment of CPS in the Program for International Student Assessment (PISA) 2015 (OECD, 2017). Besides these two studies which have been administered to large populations of students in countries around the world, there are also smaller scale studies, such as those conducted by ETS in the US (Liu, Hao, von Davier, Kyllonen, & Zapata-Rivera, 2016) and researchers (Lin, Hsiao, Chu, Chang & Chien, 2015) in Taiwan. While these reported studies differ in terms of the content focus of the problems, the collaboration arrangements, and the technology platforms used, all of these assess CPS as an individual's attribute. On the other hand, studies in CSCL generally focus on the interactions and joint activities among group members and the performance of groups and communities. Koschmann (2001, p. 19) argue that the concern of CSCL is "with the unfolding process of meaning making . . . , not so-called "learning outcomes"". Stahl (2014) further elaborated that studies of collaborative learning focusing on the outcomes of individuals misses the most important aspect and educational potential of CSCL, that the group's outcome is not equal to the aggregate of the individual outcomes, and that the group is the appropriate unit of analysis for understanding collaborative learning. Despite the above differences between CPS assessment and CSCL, this paper will present a less traditional CPS assessment study with findings which may indicate the direction of future interplay between the two domains.

## Literature review

Hesse et al. (2015, p. 38) define CPS as "approaching a problem responsively by working together and exchanging ideas". CPS thus comprises two constructs, collaboration and problem solving, as exemplified in the framework described, which is grounded on research in the CSCL area. Collaborative problem solving is similar to individual problem solving, except that in CPS all the steps are directly observable as the group members need to communicate and negotiate their understanding during each step of the process. Hesse et al's (2015) framework is adopted by the ATC21S Consortium as the assessment framework for their CPS assessment system. The PISA 2015 assessment framework (OECD, 2017) is also grounded on the same theoretical underpinnings as elaborated in Hesse et al's (2015), but expressed in a 4x3 matrix, comprising four stages of problem solving (exploring and understanding, representing and formulating, planning and executing, and monitoring and reflecting) and three communication processes (establishing and maintaining shared understanding, taking appropriate action to solve the problem, and establishing and maintaining team organization). Since in this study, the ATC21S assessment system has been deployed as one of the methods to assess students' CPS skills, we will further elaborate on this below.

## The ATC21S assessment framework

The ATC21S assessment framework has a three-level hierarchy of component skills (Hesse et al., 2015), as shown in Table 1. At the top level, CPS comprises social and cognitive process skills. The former is further differentiated into participation skills, perspective-taking skills and social regulation skills, while the latter is differentiated into task regulation skills, and learning and knowledge building skills.

Table 1: The ATC21S CPS assessment framework (Hesse et al., 2015, p. 41-52)

| Level 1                      | Level 2                                     | Level 3   |
|------------------------------|---|---|
| (1) Social process skills    | (i) Participation skills                    | (a) Action  |
|                              |   | (b) Interaction   |
|                              |   | (c) Task completion/perseverance                        |
|                              | (ii) Perspective taking skills              | (a) Adaptive responsiveness                             |
|                              |   | (b) Audience awareness (mutual modelling)               |
|                              | (iii) Social regulation skills              | (a) Negotiation   |
|                              |   | (b) Self-evaluation (Metamemory)                        |
|                              |   | (c) Transactive memory                                  |
|                              |   | (d) Responsibility initiative                           |
| (2) Cognitive process skills | (i) Task regulation skills                  | (a) Problem analysis                                    |
|                              |   | (b) Goal setting  |
|                              |   | (c) Resource management                                 |
|                              |   | (d) Flexibility and ambiguity                           |
|                              |   | (e) Information collection                              |
|                              |   | (f) Systematicity                                       |
|                              | (ii) Learning and knowledge building skills | (a) Relationships (Representations and formulations)    |
|                              |   | (b) Rules: “If...then”                                  |
|                              |   | (c) Hypothesis “what if...” (Reflection and monitoring) |

At the third or lowest level, participation skills include: (a) action – skills in performing the actions needed for the solving process, (b) interaction – skills of carrying out interactions with others in a chain of alternating actions (upon others) and in response to other’s actions, and (c) task completion/perseverance – the ability to persist in the task performance until the task completion.

Perspective taking skills include: (a) adaptive responsiveness – skills of responding to others’ views or contributions and incorporating them into one’s own thoughts and actions, and (b) audience awareness (mutual modelling) – skills of adapting one’s own views, efforts or accomplishments from others’ (or audiences’) perspectives when presenting them so that they are more easily understood or appreciated.

Social regulation skills include: (a) negotiation – resolving differences in work procedures, views, targets, priorities, potential solutions, etc. through compromise, consensus or other means; (b) self-evaluation (metamemory skill) – knowing one’s position, knowledge, strengths, weaknesses and other attributes; (c) transactive memory skills – knowing other collaborators in the same terms as in knowing oneself in metamemory skills; and (d) responsibility initiative – taking initiatives and responsibilities at different stages of the problem-solving process to achieve high-quality solutions through the joint efforts of the group.

Task regulation skills include: (a) problem analysis – dividing a problem into a set of more manageable tasks for solving the problem; (b) goal setting – setting clear, specific and achievable goals and sub-goals to direct and motivate the group’s problem solving efforts, (c) resource management – obtaining, allocating, utilizing and assessing resources based on the available pool of human resources (group members’ knowledge, skills and special capabilities) and other tangible resources such as equipment, tools and software applications, (d) flexibility and ambiguity – broadening the range of solution methods through problem representation, re-organization or new strategies, and determining acceptable tolerance of ambiguity and treating them as opportunities to further explore the problem space for alternative steps or strategies; (e) information collection – identifying and collecting information to solve a problem to meet the needs in terms of relevance, scope, detail, and time availability; and (f) systematicity – using planful, methodical approaches such as means-ends analysis and adopting results monitoring and reflection

Learning and knowledge building skills include: (a) representation and formulations – identifying relationships between elements in the problem space, and making representations and formulations of the inner connections among elements of tasks, (b) rules – ability to recognize contingencies and causal relationships as in making “if...then” rules, and (c) hypothesis – setting hypotheses by generalizing observations or parts of the solution and testing them through “what if...” inquiries, followed by monitoring and reflection of actions and



strategies.

## Design of CPS assessment systems

All CPS assessments systems are computer-based, and these can be generalized into two approaches. In the first approach, students work online on a collaborative task, communicating through a chat box, as exemplified by the ATC21S system (Awwal, Griffin, & Scalise, 2015). The CPS system licensed by the University of Melbourne for use by the research team only supports dyads (as randomly- or pre- assigned pairs) collaborating on a bundle of about three to four tasks, which are to be completed within an hour. The system is designed on the basis of a jigsaw model of collaboration such that any one person would not have enough information to complete the task without help from the collaborator. Process and click stream data generated during the task interactions are recorded in time-stamped session logs, and the chat box interactions are recorded in the chat logs. A sophisticated system of coding, mapping, and scoring was developed for each of the indicators based on the CPS assessment framework (Adams et al., 2015). The system can generate reports on individual and class (i.e. everyone who participated in the assessment in the same session) performance.

In the second approach, the PISA 2015 CPS assessment, items were constructed according to the assessment matrix such that each item can be classified as targeting one of the 12 problem-solving-task-focus and collaboration-focus combinations. Students collaborate with different numbers of computer agents and playing different roles in the interactions, according to the specific collaboration skills required (OECD, 2017). The extent to which other team members are able to collaborate can then be precisely controlled in this system. Student performance is assessed through the specific responses they select from the choices offered.

Both the ATC21S and PISA 2015 assessed CPS as a generic skill, and subject matter knowledge of the student in relation to the problem contexts is assumed to not affect the students' assessed CPS performance. A team from the US Educational Testing Service (ETS) studied students' CPS skills in the context of knowledge restructuring and revision in the study of science topics (Liu et al., 2016). They adopted the PISA 2015 CPS assessment framework in the design of their system, and conducted two studies, the first with individual students working with two computer agents, and the second with pairs of students working with two computer agents. They found that students make bigger learning gains as measured by the accuracy of their responses when they were collaborating with a human peer. However, students' CPS outcomes were not reported.

## Nature of problem and roles of team members

In the CSCL literature, knowledge building (Scardamalia, 2002) is often the most valued targeted learning outcome in the process. Thus, open-ended authentic problems closely related or meaningful to students' everyday life and requiring some sustained engagement are often selected as the targeted context for collaborative learning. This is very different from the problems used in the ATC21S, PISA 2015 or the ETS assessment systems.

Another challenge that CPS assessment systems have to confront is the roles taken up by different members in the team, beyond those captured in the collaboration aspects of the assessment framework. One particular area of concern is the role of leadership in collaborative teams, which is an important construct in the management and organizational psychology literature (Mercier, Higgins, & Da Costa, 2014). Leadership refers to the skill of taking responsibilities as well as initiatives in activities of a group to direct the group to achieve its goals (Miller, Sun, Wu, & Anderson, 2013). Leadership roles can be a result of being elected (by peers) or assigned (e.g. by teachers), but can also be emergent, that is, as a dynamic property of the situation, resulting from the social and task interactions (Gressick & Derry, 2010).

Mercier et al. (2014) further differentiated emergent leadership into two types based on the nature of the moves made by the participants, which can be focusing on organizing the group, or on advancing the intellectual aspects of the group effort. They define three types of organizational leadership moves: turn management, planning and organizing, and acknowledgement; and two types of intellectual leadership moves: idea management and development, and topic control. Mercier et al (2014) further defined emergent leadership as instances when a participant's leadership moves were taken up by others in the group. Their study included findings that have important implications for how to design CPS assessment. First, they found that the extent to which students are more willing to work on ideas from other students is affected by the collaborative medium – being greater in electronic medium like a multi-touch tablet than the paper medium, which indicates that the former is more supportive of complex conversations. Second, they also found that more than one person exhibiting emergent leadership can occur in a group. Third, by studying emergent leadership when the same groups of students engaged in tasks associated with two different subject areas, math and history, they found that the emergence of leadership differed over different task types.

## Study purpose and design

The study reported here is part of a bigger study that aims to develop an online role play serious games system that would support the learning and assessment of CPS for children aged 11 to 15. A key challenge in designing such a system is to develop valid measures for CPS assessment. In this study, our goal is to compare the CPS achievement of students as assessed by the ATC21S CPS system with a hand-coded CPS assessment based on video-recorded transactions of students working in groups over the space of two weeks to design an online game to promote cyberwellness. Their collaborative interactions were videotaped twice, with a week's separation in between. The specific research questions addressed in this study are as follows:

1. Is the ATC21S assessment framework appropriate for coding the face-to-face transactions of the students participating in the study?
2. How do students' CPS behavior and performance change over time, if at all?
3. Do the students' CPS achievement as measured by the ATC21S assessment system correlate with the hand-coded CPS scores?
4. What role can ATC21S CPS assessment scores play in the further development of CPS systems?

## Settings and participants

The study was conducted in the context of a three-session summer course for primary and secondary school students on game design to promote cyberwellness. The game design course was offered to support students interested in participating in a Minecraft coding competition. Interested students registered through the online portal of the competition organizer, which did not set any selection mechanism or criteria on the applications for participation. The summer course was held as 2-hour workshop sessions on three consecutive Friday mornings. Students were grouped into teams of three to five students of approximately the same age. The goal of the course was for the students to develop jointly a game (storyboard) that would be both interesting and help others to learn about cyberbullying and how to handle them. The students were encouraged to build their games in Minecraft, but that was not a necessary part of the course activities. A brief rundown for the three workshop sessions is as follows:

Workshop 1: Welcome, introduction to the goals of the course, introduction to examples of and ways to handle cyberbullying, group assignment, and working in groups to discuss the goals of the game they wish to construct.

Workshop 2: Introduction to game design and development, including showing examples of storyboards and Minecraft games, and group work on game design: game goal, rules, and storyboard.

Workshop 3: Group preparation on presentation of preliminary game design to other groups for peer review and feedback, further refinement and submission of final game design storyboard, and group interviews.

A total of 44 students were admitted to the workshops on a first-come-first served basis without any selection criteria. According to the University of Melbourne Assessment Research Centre, the CPS tasks have only been trialed and validated for students aged between 11 and 15, and are deemed to be too difficult for younger students. Of the 44 course participants, only 29 were aged 11 or above. They were invited to arrive at the course venue an hour before the start of Workshop 2 to take the ATC21S CPS assessment. Of these, only 14 took the ATC21S CPS assessment and gave consent to the research team to use all of the artefacts they produced during the workshops as well as all audio and video recordings for research purposes.

The group work sessions in Workshops 2 and 3 were video-recorded. Workshop 1 was not videoed as the time period available for the group work was too short. The total time of the video recording for each group was around 55 minutes in Workshop 2, and 65 minutes in Workshop 3. For the purpose of this paper, we report on one of the groups with children aged 11 or above. All four members in this group attended the three workshops and took the ATC21S CPS assessment, and the video recordings from both workshop sessions were clear.

## Results

The background of this group of students and their CPS scores in the ATC21S assessment are listed in Table 2.

Table 2: The students' demographic information and CPS achievement as assessed by the ATC21S CPS system

| Student ID | Demographic information                    | Cognitive score | Cognitive level | Social score | Social level |
|------------|--|-----------------|-----------------|--------------|--------------|
| GD28       | Female, 14 year old, 8 <sup>th</sup> grade | -0.46           | 3               | 0.85         | 5            |
| GD29       | Female, 11 year old, 6 <sup>th</sup> grade | -0.33           | 3               | 0.94         | 5            |

|      |  |       |   |      |   |
|------|--|-------|---|------|---|
| GD30 | Male, 12 year old, 6 <sup>th</sup> grade | -0.45 | 3 | 0.89 | 5 |
| GD31 | Male, 12 year old, 6 <sup>th</sup> grade | -0.19 | 3 | 1.84 | 6 |

Based on the ATC21S CPS assessment framework presented in Table 1 above, two of the research team members worked in parallel and then together to develop a detailed coding scheme for coding the video transcript. It is found that other than the off-task communications, all of the verbal transactions can be coded using the assessment framework. Other than the verbal transactions, the transcript also records instances where a member *starts* engaging in a task. These observation statements are coded under participation skills-action. An inter-rater reliability analysis on a sample of the dataset (25%) is performed to determine if there is agreement between the two independent coders. The Cohen’s kappa ( $\kappa = .82, p < .001$ ) obtained is considered to be “very good” (Alman, 1999).

As the present study is exploratory in nature, we decide to use a simple metric for our computation of CPS: that of frequency counts of turns made by each of the participants under the social and cognitive process skills categories. However, it is observed that while most of the discourse turns can be classified as instances of positive behavior in the subcategories, some of the turns are negative instances. For example, utterances such as “this is too difficult” or “I didn’t say I want to do it, I am just saying ...” are coded as *avoids undertaking the work* and is given a frequency count of “-1” to indicate that this is a negative instance of the code. Turns that are examples of ignoring other’s ideas or arguing without giving reasons are considered negative instances of social regulation.

**Table 3: The CPS scores computed based on the number of turns made by each student in the two workshops**

| Student ID             | Workshop 2 |           |           |           | Workshop 3 |           |           |           | Overall   |           |           |           |
|------------------------|------------|-----------|-----------|-----------|------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
|                        | GD 28      | GD 29     | GD 30     | GD 31     | GD 28      | GD 29     | GD 30     | GD 31     | GD 28     | GD 29     | GD 30     | GD 31     |
| Social (+ve)           | 39         | 34        | 37        | 24        | 35         | 26        | 38        | 41        | 74        | 60        | 75        | 65        |
| Cognitive (+ve)        | 22         | 9         | 10        | 9         | 8          | 9         | 16        | 11        | 30        | 18        | 26        | 20        |
| Social (-ve)           | 7          | 12        | 5         | 7         | 3          | 6         | 5         | 9         | 10        | 18        | 10        | 16        |
| Cognitive (-ve)        | 2          | 3         | 5         | 0         | 3          | 0         | 7         | 0         | 5         | 3         | 12        | 0         |
| <b>Social score</b>    | <b>32</b>  | <b>22</b> | <b>32</b> | <b>17</b> | <b>32</b>  | <b>20</b> | <b>33</b> | <b>32</b> | <b>64</b> | <b>42</b> | <b>65</b> | <b>49</b> |
| <b>Cognitive score</b> | <b>20</b>  | <b>6</b>  | <b>5</b>  | <b>9</b>  | <b>5</b>   | <b>9</b>  | <b>9</b>  | <b>11</b> | <b>25</b> | <b>15</b> | <b>14</b> | <b>20</b> |

The scores in Table 3 show a significant change in the scores associated with some of the students over the two workshops, both in terms of absolute numbers and relative to each other. The scores for GD29 and GD30 were relatively stable across the two workshops, and both increased in their cognitive scores. However, it is notable that GD29 has greatly reduced the frequencies of negative contributions in both social (from 12 to 6) and cognitive (from 3 to 0) dimensions. For GD28, while she maintained a high social score between the two workshops, her cognitive scores appeared to have been greatly reduced. On the other hand, in Workshop 3, GD31 increased significantly in his social score to rank the same with GD28 and GD30, and his cognitive score also rose to be the top score of the team. If we look at the overall scores over the two workshops, then GD28 and GD30 has the highest social score while GD28 has the highest cognitive score.

Comparing the scores in Tables 2 and 3, there is consistency in that the social scores of the students were much higher than their cognitive score. In fact, according to the ATC21S scoring rubric, 6 is the highest level of performance that a student can achieve. However, according to the scores in Table 2, GD31 has the highest score in both the social and cognitive dimensions while GD28 has the lowest scores in both the social and cognitive dimensions. How can we interpret these scores? Do they tell us anything about the students’ CPS ability at all?

### Student’s perception of their own collaboration process

Towards the end of Workshop 3, a focus group interview was held with each of the participating groups. The following were the interview questions regarding how the team worked together to develop the game:

1. How did the team come up with the game ideas presented in their storyboard?
2. How did the team go about their division of labor for the project?
3. Did the team have to look for information to complete their project, and if so, what and how?
4. Did any contradictory views arise during the process, and if so, how were these resolved?
5. Did the team encounter any problems or challenges during the work process?
6. If you are given a second chance to create a game, will you do things differently?

While these questions were designed to find out the members’ perceptions about various aspects of the

collaboration process, a surprising finding was that the role of GD28 as a leader was talked about by all the other students throughout the interview. The following is an excerpt from the part of the interview about division of labor:

*GD30: We (GD29, GD31, GD30) argued when deciding whom to be the presenter.*

*GD28: Exactly. Every time I was forced to present.*

...

*GD30: Then, we would tell her (GD28) the key points and she will be the one who present them. That's all.*

*GD31: As an analogy, we (GD29, GD31, GD30) are like the sheep while you (GD28) are the shepherd.*

Again, when asked how they would work differently if they were to create a game again (I stands for the Interviewer):

*GD28: I will run away.*

*GD29: The first thing I will do is...*

*GD31: First, I will get the team leader back. Then...*

*GD29: First I will get the team leader back and lock her up. Then...*

*I: That means you all want to listen to her instructions.*

*GD30: Yes.*

It is clear from these exchanges that GD29, GD30 and GD31 unanimously consider GD28 as their leader in the project. According to GD30, GD28 played two roles—resolving conflict and representing the team:

*GD30: The three of us often had different opinions. We would nominate our team leader whenever there was a presentation.*

## Observation of the students' collaboration process

While rigorous coding of the students' turn-by-turn transactions provide us with a theoretically grounded view of the collaboration process, it is important that we do have a qualitative, holistic sense of how the students actually worked and interacted during the two workshop sessions. It should be noted that of the four students, GD30 was the only one who did not know how to program in Minecraft. Below is a brief summary of the key episodes:

### Workshop 2 (Students were asked to work on their game design following the distributed worksheets)

1. *Deciding on the group name: GD28 led the discussion.* GD28 and GD30 constructed one using GD31's name, amid protests from GD31.
2. *Discussing game ideas.* GD28 suggested popping up questions for players to answer in Minecraft as the game mechanics. GD31 was more interested in playing around with Minecraft than completing the worksheets. GD30 requested GD31 to put aside the laptop to focus on the worksheet. GD28 then prompted the group for more game interaction ideas. GD30 put forward his ideas after hearing examples from GD28.
3. *Working on storyboard—division of labor.* GD29 asked what she should draw. After GD28 made her suggestion, GD29 expressed her desire to work in Minecraft instead. With help from GD30, GD29 was finally persuaded to work on the drawing task. GD30 then assigned a task to GD31 without giving reason, thus starting a conflict with GD31.
4. *Discussing the detailed design of the game quests.* GD28 initiated the discussion. GD29 and GD31 participated actively, but GD30 interrupted to say that this should be left to GD28 as the decision maker. GD28 paid close attention to GD29 and GD31's ideas. GD30 then join in to give his ideas.
5. *Starting to draw the storyboard.* GD28 gave suggestions to GD29 and GD31 on how to build the game in Minecraft. GD30 did not take up specific jobs but assigned jobs to GD29 and GD31.

Workshop 3 (Students were asked to give a short presentation on their game design as completed in Workshop 2 for peer feedback, and were given some time to prepare for it. After this, students were asked to work further and complete everything for a final presentation.)

1. *Discussing who should be the presenter.* GD29 and GD31 was playing on a smartphone. GD30 urged them to focus on the group discussion.
2. *Discussing the group name.* GD31 wanted to change, but no one else agreed.
3. *Discussing the detailed design of the game quests.* GD29 initiated the discussion and put forward her idea. GD30 built on her ideas and made further suggestions. GD31 disagreed with GD30 and both argued

strongly. GD28 stepped in to make other suggestions. A final agreement was reached.

4. *Creating the game in Minecraft.* GD31 expressed doubts about GD29's competence in Minecraft, ignored GD29's suggestions, turned the computer towards himself, and said that GD29 was stupid and was doing useless work. GD30 told GD31 that he was very impolite. GD29 and GD31 then entered into an intense argument over different approaches to work in Minecraft.
5. *Preparing for the first presentation.* GD28 took up the presenter role as she was considered the leader by the others. she practiced for the group presentation. GD30 asked how the game would handle the case if the player fail to achieve the game goal. GD28 and GD30 discussed the game refinement.
6. *First presentation.* Only GD28 came out to presented on behalf of the group. Not much feedback.
7. *Continuing the storyboard work.* GD28 assigned tasks to GD29 and GD31, GD30 worked with GD28 to help the other two students to complete the assigned tasks. GD29 and GD31 continued to argue, GD30 helped to clarify what needed to be done, and also helped GD31 with the task completion.
8. *Preparing for the group presentation.* GD31 finished creating part of the game in Minecraft. GD29 expressed delight and appreciation of the work by GD31. GD30 suggested to present not only the required paper-based storyboard, but also the partially completed game in Minecraft. GD31 referred to himself as the technical support for the group in building the game in Minecraft.
9. *Finalizing the game storyboards.* GD28 worked on finishing the writing part of the game storyboards. GD29 and GD30 worked together to stick pieces of the game scene paper drawings on the big cardboard for presentation.
10. *Final presentation.* GD28 was the main presenter, accompanied by the entire team. The Minecraft game was presented.

It can be seen that in Workshop 3, GD28 talked much less as she was focused on preparing for the two presentations. During this time GD30 was helping as the person to regulate the group and to resolve conflicts.

## Discussion and conclusion

We begin our discussion by addressing the four research questions of this study. First, the ATC21S assessment framework was operationalized in the hand-coding of the CPS interactions, and found to be adequate for analyzing CPS in the face-to-face context. Second, as revealed by both the CPS scores and the observations of the collaboration process, the CPS behavior and performance of individual students and the team as a whole changed over time. Third, no conclusive correlation is observed between the ATC21S assessment scores and the hand-coded scores, whether for individual workshops or aggregated over both workshops. The fourth question about how the ATC21S assessment scores may contribute to future development of CPS assessment systems is more complex, and is discussed below in terms of the issues we have uncovered in this study.

As shown in our study, even when the same assessment framework is used for the assessment, the exhibited CPS behavior (on which the score is computed) is a fluid measure, dependent on the problem context and the stage of the problem-solving task progression. GD28 may seem to have declined in her leadership role in Workshop 3 based on the scores. She had in fact been engaging much more in the completion of the worksheet task. As the group settled into a more task focused mode of operation, she did not have to spend as much time or effort on regulating the group's activities. Moreover, we can see that GD30 had been assisting her in the coordination of the project. GD30 might have been able to engage more in the task completion part of the teamwork if he had knowledge of programming in Minecraft, and may not be doing so much task and social regulation work. The particular way a student engages in a task is always dependent on the nature of the task, the familiarity of the team members with the content/disciplinary knowledge and skills required in solving the problem, their familiarity with each other, and the stage of the problem solving progress. Is it meaningful to assess the CPS skills of individuals? What might be meaningful measures of a team's CPS skills? What contextual information might need to be specified for meaningful interpretation and use of CPS measures?

From our observations of the team's work progress, the conflicts arose mainly from the need to agree on what to enter on the worksheet and how to create the game in Minecraft. Through the interactions and conflict resolution process, the students improved in their teamwork to become more productive in their collaboration. However, the scores of the team members in Workshops 2 and 3 are not able to reveal the fine grain differences and development. These scores, similar to those from the ATC21S and PISA 2015, focus only on the CPS process but not the outcomes. The primary reason for valuing CPS as a core competence is the assumption that those with higher CPS capability will delivered better, higher quality collaboration outcomes. Is it reasonable to measure CPS capability, whether as an individual or group attribute, without giving any regard to the quality of the products of the collaboration? If CPS outcome were to be assessed, how should this be done?

The four students' CPS scores have been computed by the ATC21S system, which has been developed on the basis of an enormous amount of data and shown to be reliable for the set of assessment tasks in the system. However, there is as yet no reported studies on the construct validity of these measures, i.e. how these CPS scores might inform us on the assessee's CPS ability in authentic collaboration and task settings. What might be educationally meaningful ways of using CPS assessment scores?

Our study shows that there is a strong need for more studies on assessing CPS. So far, efforts to assess CPS have been led by researchers in the psychometric and assessment communities, focusing on CPS as an individual attribute. Many of the researchers in the CSCL community see CPS as a group attribute, and outcomes are often reported within the specific study context, but not in terms of a generic individual or group attribute. We hope the issues we have uncovered will stimulate interactions between the CSCL and assessment communities.

## References

- Adams, R., Vista, A., Scoular, C., Awwal, N., Griffin, P., & Care, E. (2015). Automatic coding procedures for collaborative problem solving. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 115-132). Dordrecht: Springer.
- Altman, D. G. (1999). *Practical statistics for medical research*. New York: Chapman & Hall/CRC Press.
- Awwal, N., Griffin, P., & Scalise, S. (2015). Platforms for delivery of collaborative tasks. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 105-113). Dordrecht: Springer.
- Fiore, S. M., Graesser, A., Greiff, S., Griffin, P., Gong, B., Kyllonen, P., . . . Rothman, R. (2017). *Collaborative problem solving: considerations for the National Assessment of Educational Progress*. Washington DC: National Center for Education Statistics (NCES).
- Gressick, J., & Derry, S. J. (2010). Distributed leadership in online groups. *International Journal of Computer-Supported Collaborative Learning*, 5(2), 211-236.
- Griffin, P., & Care, E. (2015). *Assessment and teaching of 21st century skills: Methods and approach*. Dordrecht: Springer.
- Hesse, F., Care, E., Buder, J., Sassenberg, K., & Griffin, P. (2015). A framework for teachable collaborative problem solving skills. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 37-56). Dordrecht: Springer.
- Koschmann, T. (2001, December). Revisiting the paradigms of instructional technology. In *Meeting at the crossroads. Proceedings of the 18th Annual Conference of the Australian Society for Computers in Learning in Tertiary Education* (15-22).
- Lin, K. Y., Yu, K. C., Hsiao, H. S., Chu, Y. H., Chang, Y. S., & Chien, Y. H. (2015). Design of an assessment system for collaborative problem solving in STEM education. *Journal of Computers in Education*, 2(3), 301-322.
- Liu, L., Hao, J., von Davier, A. A., Kyllonen, P., & Zapata-Rivera, J. D. (2016). A tough nut to crack: Measuring collaborative problem solving. In Y. Rosen, S. Ferrara & M. Mosharraf (Eds.), *Handbook of research on technology tools for real-world skill development* (pp. 344-359). IGI Global.
- Mercier, E. M., Higgins, S. E., & Da Costa, L. (2014). Different leaders: Emergent organizational and intellectual leadership in children's collaborative learning groups. *International Journal of Computer-Supported Collaborative Learning*, 9(4), 397-432.
- Miller, B., Sun, J., Wu, X., & Anderson, R. C. (2013). Child leaders in collaborative groups. In C. E. Hmelo-Silver, C. A. Chinn, C. Chan & A. M. O'Donnell (Eds.), *The international handbook of collaborative learning* (pp. 268-279). Routledge.
- OECD. (2017). *PISA 2015 Results collaborative problem-solving*. Paris: OECD.
- Partnership for 21st Century Skills. (2009). P21 framework definitions. Retrieved from [http://www.p21.org/storage/documents/P21\\_Framework\\_Definitions.pdf](http://www.p21.org/storage/documents/P21_Framework_Definitions.pdf)
- Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Eds.), *Liberal Education in a Knowledge Society* (pp. 67-98). Chicago: Open Court.
- Singapore Ministry of Education. (2009). Third masterplan for IT in education (2009-2014). Retrieved from <https://ictconnection.moe.edu.sg/masterplan-4/our-ict-journey/masterplan-3>
- Stahl, G. (2014). The group as paradigmatic unit of analysis: The contested relationship of CSCL to the learning sciences. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.644.8856&rep=rep1&type=pdf>

## Acknowledgements

The authors wish to acknowledge that this work is funded by the Research Grants Council of the HKSAR Government, #T44-707/16/N, under the Theme Based Research Scheme.

# Metaphorical Reasoning Together: Embodied Conceptualization in a Community of Philosophical Inquiry

Sandra Lagrange-Lanaspre, LiDiLEM, Université Grenoble Alpes, sandra.lagrange-lanaspre@laposte.net

**Abstract:** In this paper, we show that taking an embodied stance at cognitive moves as not limited to speech acts but including concrete sensori-motor actions, either isolated or coverbal gestures, enriches our understanding of collaborative reasoning. More specifically, we follow the bimodal co-construction and circulation of metaphors in a Community of Philosophical Inquiry (Lipman, 2003). We show how the participants used many source concepts in the exploration of the abstract concept of ‘thinking’, and were able to build a *conceptual synthesizing metaphor*. Our results support a bimodal metaphors salience hypothesis: metaphorical gestures play a great role in making reasoning visible to others and therefore fostering collaborative learning.

## Introduction

According to Lipman, the founder of Philosophy for Children, a Community of Philosophical Inquiry (CPI), as a pedagogical practice, is supposed to favor deep collective reasoning and better cognitive understanding of a complex issue addressed through critical, creative, and caring thinking (Lipman, 2003). Its aims at teaching students how to build on opinion and to reason collectively about a philosophical question. In a CPI, students are encouraged to listen to each other with respect, to elaborate on each others’ ideas, to assist each other in justifying opinions and drawing conclusions. In Lipman’s words, it consists in “attempt[ing] to follow the inquiry where it leads rather than be[ing] penned in by the boundary lines of existing disciplines” (2003, p. 20).

But, even if the face-to-face interactions in a CPI involve multimodal communication, studies have so far mostly focused on students’ individual speech developed in such setting or on the role of the moderator described through series of verbal actions, as if learning would only take place in the brain (e. g. Simon & Tozzi, 2017). Clearly stepping away from traditional individualistic cognitivism (Menary, 2010a), adopting an embodied view on cognition offers avenues for renewing research on collective reasoning and collaborative learning. Such theoretical approach has methodological implications that guided our empirical work on a videotaped CPI among eleven 12-to-14-year-old students who are discussing about “thinking”. This study shows that taking cognitive moves as not limited to speech acts but including concrete sensori-motor actions enriches our understanding of conceptualization. More specifically, we follow the bimodal co-construction and circulation of metaphors in the CPI as activating concrete sensori-motor experience to apprehend abstract concepts collaboratively.

## Theoretical background

Our approach to metaphor is based on a strong embodied view on cognition. As it occurs in a CPI, metaphor also takes part to a more global cognitive dynamic of collective reasoning instantiating how learning is socially grounded.

### 4E cognition and the embodied approach to metaphor

Following the cognitive approach to metaphor, we consider the metaphorical process as the fundamental embodied basis of the conceptual system that makes it possible to think-and-act in our environment (Lakoff & Johnson, 1980). In this perspective, metaphor consists in building and apprehending complex abstract concepts (here after named ‘target concepts’) by comparison with familiar concrete concepts bodily experienced and transmitted (here after named ‘source concepts’). Metaphor expression exploits several modes of communication, including gesture (McNeill, 1992). In particular, Müller considers that the bimodal combination of speech and gesture is one of the key contextual resources available to catch attention and focus it on the dynamic process of metaphor activation (Müller, 2008).

More precisely, metaphorical gestures belong to the category of representational gestures, which can be defined as gestures playing a referential function by directly “*exploit[ing] imagery*” (Goldin-Meadow, 2004, p.314). A metaphorical gesture bodily refers to a concrete source concept either by drawing its shape and dimension, or by locating parts of it or tracing its trajectory or miming its actions (Colletta et al., 2010, McNeill, 1992). Such gestures are representational: they help express ideas by providing concrete images to refer to abstract concepts (e.g. the gripping and trenching gestures to express respectively founding and dividing) (Müller, 2008, McNeill, 1992).

## 4E learning and reasoning together in a CPI

### Collaborative learning as part of the social dimension of cognition

The very idea of following how metaphor are co-constructed in a CPI through a combination of speech and gesture supposes the students to deal with typical issues related to the Semin and Cacioppo's grounding Social Cognition model (2008), such as monitoring and goal-mediated synchronization. According to this model, monitoring synchronization is based on neurophysiological processes by which a perceiver mirrors actions produced by another. When subjects are engaged in a collective task, goal-mediated synchronization plays a complementary role by helping not only to mirror actions but to select adaptative and complementary responses.

Considering metaphor as an instance of collective reasoning is also consistent with studies on group cognition, which showed that taking the group as a cognitive agent helps make sense of learning processes that overcome the sum of students' individual progress (Stahl, 2006). Following an extended cognition perspective, we do not address metaphorical reasoning together as a cognitive process taking place 'in' a specific group agent, separated from its environment. We rather try to figure out how collective thinking occurs through interactions among individuals who are themselves interacting with their environment, exploiting the contextual resources available to apprehend the concepts at stake. If the group as a whole might build ownership of (part of) the metaphorical process, it is always through the environment shared among the participants. For instance, everybody might align in gesturally conferring the same specific metaphorical meaning to the same place in the room.

### Reasoning in a CPI

CPI is based on a pragmatic approach to philosophy: reasoning about abstract concepts is made possible through a confrontation of concrete individual experiences within a community. In that sense, we claim that it can be linked to embedded cognition since actions with environment are viewed as a basis for conceptualization. According to Tozzi (2009), reasoning in Philosophy for Children can be specified into three distinct reflective tasks: *problematization*, *conceptualization* and *argumentation*. *Problematization* consists in questioning our own points of view and those of others, as well as the assumptions they imply, their origin and consequences. *Conceptualization* aims at defining concepts, developing conceptual networks and distinctions. *Argumentation*, here used in a very restrictive way, consists in basing statements on rational arguments.

The more extensive view on argumentation supported by the research community on argumentation studies also offers fruitful insights about how metaphors generally contribute to reasoning. Metaphorical process is actually at the crossroads of two fundamental practices of reasoning: arguing by categorization and arguing by analogy (Plantin, 2018). In Grize's natural logic, reasoning is based essentially on the argumentative exploitation of images working as cognitive models:

Natural logic can be defined as the study of logico-discursive operations used to construct and reconstruct a schematization. The double adjective is here to highlight the fact that they are thinking operations, but only as long as they are expressed in discourse. (Grize, 1997, p. 65).

### 'Visibility' as a condition for reasoning together

Research on collaborative learning showed that just having the students work in groups is not necessarily enough to foster true sociocognitive conflict (Roschelle & Teasley, 1995) and even less unlikely to automatically results in a cognitive resolution of such conflict (e. g. Polo et al., 2016). A usual way to analyze how fruitful student-student or teacher-student interactions are in terms of learning outcomes is to specify the type of talk that they should develop (Dyke et al., 2012, Mercer, 1996). Even if most of these approaches only focus in speech, one of them draw interesting perspectives for addressing collective reasoning more globally: *exploratory talk*. *Exploratory talk* is considered as the most cognitively productive form of talk in small groups:

First it is talk in which partners present ideas as clearly and as explicitly as necessary for them to become shared and jointly evaluated. Second, it is talk in which partners reason together – problems are jointly analyzed, possible explanations are compared, joint decisions are reached. From an observer's point of view, their reasoning is visible in the talk. (Mercer, op. cit., p. 363)

Interestingly, a key property of *exploratory talk* is to make ideas and arguments *visible* to others. It seems that real-time 'visibility' of thinking-in-action is an essential condition for student-student interactions to actually foster collective reasoning. We assume that such 'visibility' can be reached and studied with a multimodal approach, notably through metaphorical gesture.



## Methodological approach to an empirical study of CPI

Our approach is deeply empirical. We provide a qualitative multi-level case study describing how bimodal metaphorical reasoning takes part to collaborative learning in a session of CPI. After specifying the pedagogical situation studied and the data collection, we present our research questions and working hypotheses. This methodological section ends with an explanation of each of our analytical steps.

### Pedagogical situation and data collection

The analyzed situation occurred in the particular context of a CPI demonstration. The session was moderated by M. Sasseville, one of the leading expert in the field, and observed by teachers, researchers, CPI practitioners and parents. It took place in November 2015, in an urban lower-class French secondary school, and involved a group of eleven students familiar with Philosophy for Children, aged 12 to 14 years, who volunteered to come on a Wednesday afternoon, a time when the school is usually closed. Participants sat in an arc facing observers who were placed on bleachers. The philosophical issue discussed was the definition of thinking. M. Sasseville started the CPI with the prompting question: "Where do thoughts come from?". The whole CPI was videotaped using a 360° camera situated in the center of the participants' arc. Speech was fully transcribed using Elan® software. Gestures were only annotated in selected episodes especially relevant for analysis.

### Research questions and working hypotheses

We explored the role of bimodal metaphors in collective reasoning in the context of this CPI. This global purpose consisted of four research questions:

1. How do the final analogies used by the students apprehend the *target concept* of thinking?
2. Does the metaphorical process trajectory evidence collective elaboration of metaphors?
3. Is metaphorical thinking together associated with advancement in reasoning?
4. More specifically, how do metaphorical gestures take part to conceptual complexification?

We based our work on the two following fundamental hypotheses:

1. Metaphors emerge and evolve through social interaction and interaction with the environment, during a CPI, describing conceptual trajectories that we can retrace *a posteriori*.
2. The salience of co-verbal metaphorical gesture make the associated cognitive process more visible and therefore foster the collective appropriation of the metaphor.

### Analytical steps: Following conceptual co-construction through bimodal metaphors

Our analytical strategy was to follow how conceptual co-construction occurred through metaphors using both verbal and gestural thinking and communicative resources. We proceeded in two main analytical steps. First, we coded the whole CPI discussion using Elan software, in two different ways, in order to identify episodes of special interest. We then undertook a multi-level qualitative case study based on some of these episodes.

#### Systematic coding with Elan® software: coupling two complementary approaches

Two complementary coding schemes were held separately to identify episodes of special interest in Elan®, the first focusing on *source concepts* and the second on *cognitive models*.

The *source concept* coding scheme, applied to all the discussion preceding the closing sequence, consisted in 5 main annotating lines. The first line was used to identify and chronologically number episodes in which *source concepts* were used. Lines 2 and 3 were used to systematically annotate respectively verbal and gestural *source concepts*. For each of these lines, two dependent lines were added to relate each expressed *source concept* to a *target concept* and to a speaker. This systematic coding allowed us to highlight *source concepts* flowing and to calculate the frequency of each *source concept*.

The *cognitive model* grid was based on an exploratory coding of the closing sequence, which was later on applied to the rest of the discussion. The analysis of the final round table empirically *grounded* the construction of this second coding scheme (Charmaz, 2006). The analytical grain used was larger, corresponding to the level of the analogies made by the students to describe the main target concept: each code related to a metaphorical *cognitive model* helping them define 'thinking'. Such comparisons are explicit analogies in the closing sequence, but they may be referred to in ways that are more implicit at different stages of the metaphorical trajectory retraced during the preceding discussion. Actually, our aim was to go back in time in order to figure out how the *cognitive model* underlying each final explicit analogy was gradually constructed through the CPI. In the second coding step, we applied the code of a *cognitive model* whenever at least one key defining feature of the *model* was referred to, either gesturally or verbally, during the discussion. Obviously, such previous occurrences of the *models* were

generally less complex than the final analogies, which is consistent with our hypothesis 1 stating that the metaphor gradually shapes through a specific cognitive trajectory.

### From the whole discussion to the conceptualization phases: multi-level deep qualitative case study

Crossing the results from the two independent coding of *source concepts* and *cognitive models*, we were able to identify that one particular *cognitive model*, ‘using a file’, was emblematic of the process of metaphorical collective thinking that we were interested in. We decided to focus on this case to conduct a multi-level deep qualitative study instantiating bimodal metaphorical reasoning together. We first specified how it helped the students apprehend the *target concept* of thinking in comparison with the other *cognitive models* that they used. Such analysis, coupled with a visualization of the chronological occurrences of the different *cognitive models* along the discussion, then allowed us to characterize how it related to the other models. After addressing the macroscopic level of reasoning progress through alternation and co-enrichment of multiple metaphors, and the mesoscopic level of the cognitive features composing this specific *cognitive model*, we zoom in to the micro-analysis of the bimodal complexification of such model through three conceptualization phases.

### **Results: bimodal construction of metaphor as a ground for reasoning together**

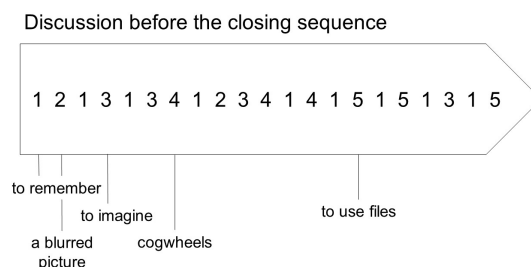
In this section, we present the results of this deep qualitative case study. First, the macro-level and meso-level analyses show how a specific *cognitive model* plays the role of a *conceptual synthesizing metaphor* (CSM). We then specify the micro-level bimodal collective complexification of such CSM.

### Metaphorical thinking together: macroscopic and mesoscopic analyses

#### Final analogies as *cognitive models* of ‘thinking’

During the closing sequence, when the students are asked by the moderator to conclude by making an analogy for ‘thinking’, they explicitly mention 9 cognitive models of this target concept, which were identified through an exploratory coding (here after mentioned in the chronological order of emergence in the conversation, in students’ own words): *to dream, a milky way, to use files, cogwheels, to reflect, to remember, a cloud, a blurred picture, to imagine*. We used all the verbal and gestural information provided by the participants to characterize these analogies as cognitive models, in order to address our first research question about how they allowed the students to apprehend the abstract concept of ‘thinking’. At most, this bimodal semantic analysis allowed for characterizing the 5 following aspects: nature of the fundamental units composing the system; whether it corresponds to a dynamic process or a static state; internal structure; localization; size (Polo, in press). In this closing sequence, some *cognitive models* are only mentioned without being described along these 5 characteristics, for other *models*, only 2 or 3 of these aspects are specified. Sometimes, the information is only conveyed in gesture, as, for instance, the localization of dreaming “in the head”, or the dynamic processual dimension of cogwheels.

Such characterization of the 9 *cognitive models* reveals that all the students, at the end of the CPI, share a few common features in their apprehension of what ‘thinking’ is. First, it is a system consisting of fundamental units, the thoughts, which can be compared to pictures. Second, it involves chronological or causal relations that allow switching from one thought to another. Third, thinking is presented as an internal process localized within the cognizer. It is nothing but surprising that the students share some conceptual features about thinking at this moment of the CPI, since they have been discussing about this concept for more than 40 minutes.



**Figure 1.** Emergence and reuse of 5 *cognitive models* of thinking along the CPI.

Nevertheless, coding of the *cognitive models* along the preceding discussion revealed that only 5 of them were used in the CPI before the closing sequence: *to remember, a blurred picture, to imagine, cogwheels* and *to*

*use files*. Still, none of them was used neither only once nor by only one person. Such result both confirms our first hypothesis and evidence collective elaboration of metaphors (research question 2). As figure 1 shows, during the whole discussion preceding the closing sequence, the students experienced 21 episodes of focus on one single *model* (9 on *to remember*, 2 on *a blurred picture*, 3 on *to imagine*, 3 on *cogwheels*, and 3 on *to use files*). The co-construction of metaphors supporting these *cognitive models* was not linear: the students did not fully elaborate one *model* before considering another one, but rather tended to switch from one *model* to another in a heuristic way. Contrasting different metaphorical views and their ability to apprehend the target concept, such alternation may be a pattern necessary to foster the deepening of each *cognitive model*.

**Metaphorical dialectical opposition between two models: *to remember* versus *to imagine***

Such heuristic process is made visible through a bimodal analysis of each of these episodes. We created 5 collections of episodes focusing on the same *cognitive model* and paid a special attention to the corresponding metaphorical gestures. Such analysis revealed strong specificities associated to each *model*:

1. *To remember*: pictures from the outside are coming up to the head (44 s collection)
2. *A blurred picture*: thinking is a flow of blurred pictures (10 s collection)
3. *To imagine*: pictures from the inside are coming out of the head (15 s collection)
4. *Cogwheels*: each thought is a cogwheel which move causes another cogwheel’s move (17 s collection)
5. *To use files*: we can create files, open them, modify them, delete them, retrieve them (53 s collection).

This gestural description of the models reveals that a strong binary opposition was co-constructed along the discussion between to think as *to remember* or as *to imagine*. The expert moderator himself makes this opposition explicit and gesturally expresses it three times during the discussion, by the 15<sup>th</sup>, 18<sup>th</sup> and 28<sup>th</sup> minute. Our interpretation is that confronting these to metaphors worked as a dialectical tool to gradually sophisticate students’ understanding of the *target concept*. Such advancement in reasoning also proceeded through metaphorical thinking, by elaborating what we called a ‘conceptual synthetizing metaphor’ (CSM) that overcame such opposition and conciliated the strength of several *cognitive models* to better apprehend the *target concept*.

***To use files* emerging as a conceptual synthetizing metaphor (CSM)**

The cognitive model *to use files* is the latest to emerge (from episode 15<sup>th</sup> onwards), and is co-elaborated by Nourra and Jean-Luc. Finally, Jean-Luc reuses it as a full complex analogy during the closing sequence. For its late emergence in the discussion and its ability to conciliate several key features previously introduced in other metaphors, this *cognitive model* plays the role of a *conceptual synthetizing metaphor* (CSM). With respects to our third research question, a CSM is defined as a metaphor associated to advancement in reasoning by encompassing several aspects of the *target concept* that could so far only be apprehended by multiple metaphors. In this case, *to use files* is clearly a CSM allowing the participants to overcome the opposition *to remember* vs *to imagine*.

Table 1: Features of the three *cognitive models* ‘to remember’, ‘to imagine’, and ‘to use files’ (final round table)

|                     | Fundamental unit  | Dynamic process ?                         | Internal structure | Location | Size |
|---------------------|---|---|--------------------|----------|------|
| <b>To remember</b>  | Evan <i>sometimes at thinking one is uh: like remembering</i><br>Nourra <i>take an object, a place and.: recording uh: this place this object in our head and: remembering it but uh: [...] not as clear as in the real world</i> |   |                    |          |      |
|                     | Pictures<br>- of objects<br>- of places   | Registering<br>Reminding                  | Blurred aspect     | Head     |      |
| <b>To imagine</b>   | Sofia <i>thinking is like [...] imagining something</i>   |   |                    |          |      |
|                     | Pictures  | Active cognizer<br>creating pictures      |                    |          |      |
| <b>To use files</b> | Jean-Luc <i>[...] it’s a movement in a huge archive where some files are created sometimes by chance mixing some up [...] thinking is [...] moving uh: // using the files that we have in our head</i>                            |   |                    |          |      |
|                     | Files created<br>- randomly<br>- from previous ones   | Creation<br>Displacing<br>Intentional use | Chronology         | Head     | Big  |

Table 1 summarizes key characteristics of the latter three *cognitive models* that appeared during the closing sequence. Both *to remember* and *to imagine* are described verbally and gesturally as dynamic processes based on pictures. *Remembering* is depicted as a double process taking place in the head, consisting in a first step in registration of pictures from the outside, and in a second step in reminding them, which is to say accessing these pictures within brain memory. *Imagining*, in contrast, depict a more active subject, constructing and sharing mental pictures created in his head. The CSM *to use files* is the only final analogy fully described into all the 5 parameters of size, localization, structure, degree of dynamism, nature of fundamental units. In this model, the fundamental units of thoughts are depicted as files, which can either be created ‘by chance’ or from a ‘preexisting file’. As a whole, the archive room located in our head is big and well chronologically structured. It hosts either intentional or random processes of file creation, displacing and deletion. We understand the idea of random processes encompassed by this CSM as an integration of the idea that thoughts can come from pictures from the outside randomly registered and reminded – a key aspect of the cognitive model *to remember*. Nevertheless, such CSM also includes the idea of intentionality that was brought from the model *to imagine*. The image that one file might be created from a previous one is a way to conciliate the idea of subject agency with the consideration that a thought is not created from scratch; such causal link can also be understood as integrating the *cogwheels* model.

### Conceptualization: bimodal collective complexification of the CSM at the micro level

Applying a micro-analysis based on the source concept coding scheme presented above, we then show that the CSM to use files first emerged through an extended bimodal metaphor about forgetting and remembering. This metaphor was collectively built through three main phases of complexification: “metaphor introduction”, “metaphor specification” and “metaphor exploitation”, corresponding to three distinct stages of conceptualization.

#### Metaphor introduction: *forgetting* is like a file that “we don’t open”

Figure 2 below corresponds to the verbal transcript and the key gesture of the metaphor introduction phase.

---

367 : Nourra : (...) about forgetting it’s quite like a file that (...) we keep except that::: // [1] {Ulrick: we don’t open it} / yes that’s it we’ll open it later (...)




---

Figure 2. Metaphor introduction phase transcript and illustration of gesture 1.

During this phase, Nourra first formulated a verbal comparison for defining the target concept of *forgetting*. Then Nourra added a gesture to her speech (see illustration), which expressed the *source concept* ‘to open a file’. While this gesture was synchronous with the words “except that:::”, which conveyed Nourra’s hesitation, Ulrick immediately suggested a verbal formulation for it: “we do not open it”. A bimodal metaphor was co-constructed about the notion of *forgetting*, and our view is that the process of monitoring synchronization described by Semin and Cacioppo (2008) allowed Ulrick experience the gestural *source concept* represented by Nourra and helped him access to the associated verbal concept. In addition, we note that the moderator illustrated the child’s speech with one gesture that expressed the *source concept* ‘to put a file aside’. We also note that the child imitated this gesture, immediately and many other times, which can also be explained as a monitoring synchronization process. On the reflective plane, Nourra was defining the concept of *forgetting*, exemplifying Tozzi’s *conceptualization*.

#### Metaphor specification and conceptual distinction: *remembering* is like “opening the file”

Figure 2 presents the verbal transcript and illustration of four key gestures of the metaphor specification phase.

---

379 : Nourra: Well, remembering it’s:: it/ what I was talking about // here it’s opening the file // [2a, moderator gesture]  
 380: moderator: rather than putting it aside [2b] it’s {Nourra : yeah} to open it  
 381: Nourra: it was [2c] put aside and [2d] now we open it



Figure 3. Metaphor specification phase verbal transcript and illustration of gestures 2a, 2b, 2c et 2d.

During this phase, Nourra created an extended the metaphor previously introduced. She used the same verbal *source concept* ('to open a file') as for defining the notion of *forgetting*, but this time for specifying the opposite concept of *remembering*. Even though the reflexive task was still conceptualization, it was now more specifically a task of distinguishing concepts. Subsequently, our coding scheme allowed us to reveal a collective synthesis. The moderator synthesized and rephrased the conceptual distinction between *forgetting* and *remembering* through a sequence including two bimodal metaphors. In the first bimodal metaphor, he repeated his gesture done during the metaphor introduction (see illustration of gesture 2b above) but adding the words "rather than putting it aside". This first bimodal metaphor was opposed to a second one, in which he created a palm up gesture representing the *source concept* of *openness* to the words "we open it". Again illustrating the monitoring synchronization process mentioned above, Nourra immediately validated this bimodal synthesis of conceptual distinction made by the moderator. She said: "it (the file) was put aside but now we're opening it", while doing gestures similar to those of the moderator, conveying the *source concepts* 'to put a file aside' and 'to integrate something'. Hence, during this phase, the student and the facilitator through verbal and gestural modalities were gradually and collectively constructing a bimodal metaphor.

#### Metaphor exploitation and conceptual network elaboration: moving files in a big archive

Finally, in speech turn 391, another child used the file extended metaphor for creating a complex conceptual network. Figure 4 below shows the verbal transcript and two key gestures of this metaphor exploitation phase.

391: Jean-Luc: (...) actually according to me remembering well I'm going to cite this example again it's like // memories are like a huge archive and when you think you move from one file [3a] to another and therefore // heum sometimes when you try to think about something // you can't open this file and / and this is when you forget something and sometimes // uh when you succeed in remembering something you succeed in opening the file and this is because you have moved from one file [3b] to another



Figure 4. Metaphor exploitation phase verbal transcript and illustration of gestures 3a and 3b.

During this third phase, Ulrich created a new bimodal synthesis of the conceptual distinction between *forgetting* and *remembering* by taking up speech and gestures of his classmate. He added two similar bimodal metaphors through which *thought* was apprehended in the terms of the *source concept* 'moving from one file to another'. In addition, he also expressed the target concept of *memory* successively through the gestural *source concept* of 'cycle' and the verbal one of 'archive'. Jean-Luc created a conceptual network connecting the concepts of *forgetting*, *remembering*, *memory* and *thought* by taking up Nourra's relevant speech and gestures, and adding complementary bimodal elements in order to contribute to the collective advancement in conceptualization of the notion of *forgetting* and *remembering*. This result could illustrate the complementarity of monitoring synchronization and goal-mediated synchronization implied in the Semin and Cacioppo's Social Cognition model.

## Discussion and concluding remarks

Studying the bimodal co-construction of metaphors along a CPI reveals how strongly embodied metaphorical reasoning plays a role in collaborative learning and conceptual change. Using the resources of their environment

and the interactions with each other, the participants used many source concepts in their exploration of the abstract concept of ‘thinking’, and were able to build heuristically a *conceptual synthesizing metaphor* through the *cognitive model* ‘to use files’. The feature of the model that they used the most throughout the bimodal construction of this metaphor was ‘to open a file’, which corresponds to the only feature that was expressed both verbally, gesturally and collectively from its introduction phase onwards. This result supports and complete the bimodal metaphors salience hypothesis: metaphorical gestures, especially if produced by a hearer and not the speaker, plays a great role in making reasoning visible and therefore fostering collaborative learning. Further research might include multi-level studies of bimodal metaphorical reasoning in other collaborative learning contexts. Gathering more data with, when relevant, statistical treatment on the frequency and re-use of some metaphorical features may result in a theoretical generic modelisation of metaphorical reasoning together. For now, practical implications of this exploratory study still need to be addressed: how can the CSCL community seriously take into account the strongly embodied nature of metaphorical reasoning together at designing online educational interactions? A key might be to paid attention to the contextual resources offered by such environments for embodied metaphorical thinking.

## References

- Charmaz, K. (2006). *Constructing grounded theory: a practical guide through qualitative analysis*. London: Thousand Oaks, Calif.
- Colletta, J.-M., Capirci, O., Cristilli, C., Goldin-Meadow, S., Guidetti, M. & Levine, S. (2010). Multimodal Data Transcription and Annotation with ELAN. Coding Manual, ANR-05-BLANC-0178-01 & -02, Lidilem, University of Grenoble Alpes.
- Dyke, G., Adamson, D., Howley, I., & Rosé, P. C. (2012). Towards Academically Productive Talk Supported by Conversational Agents. In S. Cerri, W. Clancey, G. Papadourakis, & K. Panourgia (Eds.), *Intelligent tutoring systems*, 7315 (pp. 531-540). Springer.
- Goldin-Meadow, S. (2004). *Hearing gesture: How our hands help us think*. Cambridge: Harvard University Press.
- Grize, J. B. (1997 [1990]). *Logique et langage*. Ophrys.
- Lakoff, G., Johnson, M. (1980). *Metaphors We Live By*. University of Chicago Press.
- Lipman, M. (2003). *Thinking in Education*. Cambridge University Press (2<sup>nd</sup> Edition).
- McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. University of Chicago Press.
- Menary, R., (2010). Introduction to the special issue on 4E cognition. *Phenomenology and the Cognitive Sciences*, 9, 459-463.
- Mercer, N. (1996). The quality of talk in children’s collaborative activity in the Classroom. *Learning and instruction*, 6(4), 359–377.
- Müller, C. (2008). What gestures reveal about the nature of metaphor. In A. Cienki & C. Müller (Eds.), *Metaphor and Gesture* (pp. 219-247). Amsterdam/ Philadelphia : John Benjamins Publishing Company.
- Plantin, C. (2018). *Dictionary of Argumentation*. London: College Publications.
- Polo, C. (in press). La participation des gestes métaphoriques à la construction collective d’analogies cognitives en CRP. In E. Auriac-Slusarczyk (Ed.) *Les ateliers de philosophie. Une pensée collective en marche*. Sphères Educatives.
- Polo, C., Lund, K., Plantin, C., Niccolai, G. P. (2016). Group Emotions: The Social and Cognitive Functions of Emotions in Argumentation. *International Journal of CSCL* 11(2), 123-156.
- Roschelle, J., & Teasley, S. D. (1995). The construction of shared knowledge in collaborative problem solving. Proceedings of CSCL (Bloomington, IN). In C. O’Malley (Ed.) *Computer-Supported Collaborative Learning* (pp. 69–97). Berlin: Springer Verlag.
- Semin, G.R. & Cacioppo, J. (2008). Grounding social cognition: Synchronization, coordination, and co-regulation. In Semin, G.R. & Smith E.R. (Eds.), *Embodied grounding: Social, cognitive, affective, and neuroscientific approaches* (pp. 119-148). Cambridge University Press.
- Simon, J.-P. & Tozzi, M. (Eds.) (2017). *Paroles de philosophes en herbe*. Grenoble: UGA Editions.
- Stahl, G. (2006). *Group cognition*. Cambridge: MIT Press.
- Tozzi, M. (2009). Apprendre à penser ensemble. *Diotime*, 42.

# Tracing Teacher Collaborative Learning and Innovation Adoption: A Case Study in an Inquiry Learning Platform

María Jesús Rodríguez-Triana, École Polytechnique Fédérale de Lausanne and Tallinn University, mjrt@tlu.ee

Luis P. Prieto, Tallinn University, lprieto@tlu.ee

Tobias Ley, Tallinn University, tley@tlu.ee

Ton de Jong, University of Twente, a.j.m.dejong@utwente.nl

Denis Gillet, École Polytechnique Fédérale de Lausanne, denis.gillet@epfl.ch

**Abstract:** Social processes play an important role in teachers' ongoing professional learning: through interactions with peers and experts to solve problems or co-create materials, teachers internalize knowledge developed in their communities. However, these social processes and their influence on teachers' learning (i.e., adopting new practices) are notoriously hard to study, given their implicit and informal nature. In this paper, we apply the Knowledge Appropriation Model (KAM) to trace how different social processes relate with the implementation of pedagogical innovations in the classroom (as a marker of professional learning), through the analysis of more than 20,000 artifacts from Go-Lab, an online community to support inquiry-based learning. Our results not only show how different social processes like sharing or co-creation seem to be related to increased classroom implementation. Also, they provide insights on how we can use traces from digital co-creation platforms, to better understand the social dimension of professional learning.

## Introduction

Despite the benefits that educational innovations may entail, they require the adoption of new teaching practices (i.e., a form of professional learning), which is known to be challenging (Webb & Cox, 2004). The adoption of innovations among professionals builds on processes of knowledge creation that span individual, group and collective levels (Nonaka, von Krogh, & Voelpel, 2006): interacting with peers, seeking help from more expert colleagues, etc. Indeed, communities of practice (Wenger, 1999) have a crucial role on workplace learning, as professionals get access, adapt and internalize knowledge that has been developed in the community (often reified as documents, materials, lesson plans and other artifacts).

From a research standpoint, however, a challenge in studying this kind of professional learning is that many of these processes and knowledge are informal, implicit, and hard to trace (often happening face-to-face, across multiple spaces and timeframes). One opportunity to do so is the use of digital platforms to embody and facilitate many of these social processes. In the concrete case of teacher professional development, the use of co-creation platforms to develop materials has a long history (e.g., Vuorikari et al., 2011). However, the fact that professional development actions (e.g., training workshops) still remain quite separate from classroom implementation of the target innovations (which is the hallmark of teachers actually learning to use the innovation professionally), makes tracing this kind of complex ongoing professional learning rather elusive.

Recent advancements in research about professional and workplace learning aim at clarifying this complex system of social processes and practices. For instance, the Knowledge Appropriation Model (KAM) (1) identifies different kinds of knowledge maturation, scaffolding and appropriation practices at different social levels, as crucial in such learning process.

In this paper, we explore how the relationship between such knowledge appropriation practices (which happen at the individual, group and community levels) and the adoption of educational innovations (as a marker of professional learning) can be traced, through the study of Go-Lab, an online community aimed at supporting teaching and inquiry-based learning. Our ultimate goal is to better understand which social processes seem to be more strongly related with actual classroom implementation, to design interventions and further develop such platforms to encourage teacher adoption and learning.

An important aspect that enables this kind of study is the fact that Go-Lab's ecosystem supports not only the co-creation (with peers or experts) and sharing of materials and plans reifying the inquiry learning pedagogy (called Inquiry Learning Spaces, or ILSs). It also supports the implementation of these lessons with students in the classroom (hence making innovation adoption somewhat traceable). Hence, in this paper we trace the social practices defined in KAM as these ILSs are co-created, shared, etc. Our analysis of 24,485 such learning artifacts, created between November 2015 and November 2018, through descriptive and exploratory analyses, try to answer the following research question: *How are the knowledge appropriation collaborative practices related with the teacher adoption of inquiry-based pedagogy in the Go-Lab community?*

The rest of the paper is structured as follows: the next section summarizes related research on knowledge appropriation and professional learning theories. Then, we describe the methodology we followed in our study and, later on, the context (including how we operationalized the KAM in Go-Lab) and the results of our study. We end the paper with implications and future work.

## **The role of social learning and knowledge appropriation in innovation adoption**

In learning technologies research and many adjacent fields (like CSCL), the adoption of educational innovations is often studied in terms of acceptance models, such as TAM (technology acceptance model - see, e.g., Teo, 2009). This kind of models assume that the adoption of innovations in teaching is influenced by teacher's competence, self-efficacy and perceived ease of use and usefulness of the innovation. However, these models treat "technology" (or whatever other innovation) as an object that either is -or is not- appropriated. As such, they do not take into account processes of co-creation in which new practices emerge. In contrast, sociocultural learning theories look at innovation adoption as a form of social learning that is connecting individual learning to the emergence of common practices and the development of knowledge in groups and communities. Examples of this view include Triological Learning (Paavola & Hakkarainen, 2014) or Knowledge Maturation (Maier & Schmidt, 2015) models.

Indeed, among the many factors for the adoption of pedagogical innovations, social learning has been shown to have an important, positive impact: the more teachers are involved in professional networking and collaboration, the more likely they are to apply such pedagogies (e.g., OECD, 2014). Similarly, teacher collaboration and efficacy are correlated, and seem to predict student achievement (e.g., Lee, Dedrick, & Smith, 1991). Moreover, different kinds of collaboration have also been shown as important: understanding develops and new classroom practices emerge in the context of teachers' collaboration with peers (Blumenfeld, Krajcik, Marx, & Soloway, 1994), but also with experts (e.g., Schenke, van Driel, Geijsel, Sligte, & Volman, 2016).

It is however an open question "how such collaborative processes can best occur, what makes them effective and what are the limits of their efficacy" (Billett & Choy, 2013, p. 268). The recently proposed Knowledge Appropriation Model (KAM) (1) describes those collaborative practices that are important in the context of adopting innovations. The model identifies informal learning practices (in organizations, communities, groups and individuals) that can be observed in the context of innovation adoption. The model is based on existing sociocultural models of learning (knowledge maturation and scaffolding) and explains how these processes are interconnected in workplace learning.

In the context of KAM, knowledge maturation is understood as a social learning process that transforms knowledge (often embodied as artifacts) from the individual level into communities of interaction (by participating in activities of social groups or communities) (Maier & Schmidt, 2015). Scaffolding describes a related social learning process whereby individuals develop expertise through guided experience with experts or more advanced peers who help them to internalize knowledge that has been developed.

In previous research on workplace learning, 12 collaborative learning practices have been derived, and are assumed responsible for the successful adoption of an innovation. In the context of teacher professional learning, they would be:

- *Knowledge maturation* practices lead to the transformation and maturation of knowledge (Figure 1, left). Starting from a teacher who takes up materials for new teaching and learning methods (appropriating an idea), this knowledge is made accessible to a group of people (sharing) and further transformed (co-creation) into more mature knowledge so that they are reusable for other teachers outside the narrower community that has created it (formalization). Eventually, knowledge might reach a status in which it becomes a standard, e.g., in terms of national curricula or entering into widely accepted training material for beginning teachers (standardization).
- *Knowledge scaffolding* practices explain how this knowledge is applied in concrete working situations (Figure 1, right). A teacher may request help regarding a certain problem, and peers or experts provide guidance towards a solution. Then, as the teacher acquires competence, peers and experts fade support.
- *Knowledge appropriation* practices (Figure 1, center) ensure that the adoption of innovations is successful, sustained and scaled. These practices describe how individuals are made aware of knowledge about typical problems in the domain and possible solutions (create awareness), and how the community maintains maintains a shared understanding about these problems and the solution. Later, teachers can adapt those solutions to new situations according to the local circumstances, and establish some form of validation (e.g., gathering experiences or formal evidence about the success and impact of the solution). Knowledge appropriation practices are assumed to underlie both maturation



and scaffolding, and these and related practices have widely been discussed in the collaborative learning literature to be conducive to social learning.

As collaborative online platforms are increasingly being used in the collaborative design of innovative teaching and learning scenarios, it is possible to find evidence of these practices in those online platforms, and to gain an understanding of their role in the process of adopting an innovation. These insights may also help us suggest ways to improve such collaborative platforms.

In this paper, we use KAM as an analytical framework to guide the analysis of how knowledge appropriation practices and adoption of an educational innovation (as a marker of collaborative professional learning) can be traced in a (large-scale) collaborative online platform. The following section shows how the KAM helped us describe the case of the Go-Lab online community (to support IBL) and identify indicators to follow these practices and the adoption of IBL.

## **Context: The Go-Lab ecosystem to support Inquiry-Based Learning**

Inquiry-based learning (IBL) has long been associated with positive student outcomes like an improved ability to apply scientific thinking, or increased retention rates (Seymour, Hunter, Laursen, & DeAntoni, 2004). However, it has proven a major challenge to get large numbers of teachers to use it (Fairweather, 2008; Henderson & Dancy, 2011): since effective support needs to be offered to students in the process of inquiry (de Jong, Linn, & Zacharia, 2013; Kirschner, Sweller, & Clark, 2006), it is a notoriously demanding practice for teachers. In that sense, the support to the design, or creation, of IBL activities is one of the most-often cited (e.g. Slotta, Tissenbaum, & Lui, 2013).

The Go-Lab initiative aims to support teachers in the adoption of IBL pedagogies and on-line labs, by offering a technological ecosystem around the notion of Inquiry Learning Spaces (ILSs): pedagogically-structured learning environments that can contain labs, apps and resources (Rodríguez-Triana et al., 2014). Typically, an ILS is visualized as a web space with a set of folders (or tabs, in the student view) for each of the phases of an inquiry (from hypothesis formulation to data interpretation). These sub-spaces then contain different apps, online labs or other learning resources for students to perform the inquiry.

The Go-Lab ecosystem is composed of two platforms: a) Golabz (2), a repository where teachers can find ready-to-use apps, labs, ILSs, and support materials; and b) Graasp (3), a platform for ILS authoring and monitoring, and community gathering. Additionally, there is a help desk where teachers can request support, and specific face-to-face training events that are organized at the regional, national and international level.

To understand how the Go-Lab ecosystem mediates teachers' social learning of the inquiry-based pedagogy, let us look at a typical example of usage. After a teacher has become acquainted with inquiry-based learning (e.g., in an online or face-to-face training event, or on her own), the teacher may decide to put it in practice in her classroom. She can, for instance, browse, find and *reuse* one of the ILSs available in the Golabz repository. If she decides to *create* the lesson from scratch, or needs to *adapt* the original ILS, she is directed to Graasp, where the ILS can be modified. At any point in this process, the teacher can *share* her work-in-progress lesson with other teachers or experts in the domain, who can *co-create* the ILS through Graasp as well. Communication during this co-creation can happen face-to-face, or through Graasp's built-in chat (associated to each ILS). When the teacher is satisfied with the ILS and deems it ready for use in the classroom, she can direct her students to the *student-view* of the ILS: a simpler interface following the structure designed by the teacher, where students can follow the activities, input their hypotheses, data, conclusions, etc. Later on, the teacher(s) can create further *copies* of the ILS (e.g., for use with other student groups of hers). Eventually, if the teacher thinks the lesson design may be useful for other teachers as well, she can submit her ILS for *publication* in the Golabz repository, for use by the larger teacher community. During this publication process, experts will curate the ILSs, collaborating with the teacher, suggesting further enhancements or modifications.

Go-Lab has been the object of several EU projects, and as of November 2018, it has reached more than 26,000 teachers and 73,000 students all over the world, receiving on average more than 300 visits per day. While teachers can use isolated apps or labs, 63.11% have been involved in the (co)creation of more 37,000 ILSs. However, according to the number of "student-views", only 5.51% of these ILS may have been implemented in the classroom.

It would thus be interesting to understand what social processes seem to be most associated with classroom implementation as a signal for the effectiveness of the collaborative professional learning of teachers involved in the platform. Or, said differently: *what are the social practices most followed by Go-Lab's classroom adopters?* In the study section below, we detail how we have applied the KAM to guide the identification of metrics about teacher's collaborative learning practices in Go-Lab, and explore how teachers' ILSs created in different (collaborative) ways are associated with their classroom implementation – as a marker of the teachers actually learning the IBL pedagogy they represent.

## Research questions, hypotheses, and methodology

Against the aforementioned research background and context, this paper tries thus to answer the following research question: *To what extent are collaborative professional learning practices (as defined by KAM) related to teacher adoption of IBL pedagogy in Go-Lab?* According to KAM, two hypotheses can be derived from it:

- *Hypothesis 1: the ratio of adopted ILSs will increase across the categories of maturation*
- *Hypothesis 2: the ratio of adopted ILSs will be higher for ILSs that had been reused than for those that had not been reused*

In order to validate these hypotheses, we have looked at the data available in the Graasp database and logs. In our study below, we use the ILS as the main unit of analysis, as it is the smallest entity in the system that has a pedagogical meaning, and on which knowledge appropriation actions (like sharing or co-creation) can be performed. More concretely, we have analyzed the data about the ILSs created from November 2015 (when authoring and implementation actions started to be tracked in the platform) to November 2018. Since we are interested in tracing the collaborative practices of teachers, only teacher-created spaces have been considered in the analysis (i.e., ILSs created by experts or other project members have been discarded).

Several assumptions have been made to interpret the data available in the logs. For the purposes of the description of results and discussion that follow, we considered as *experts* those registered users that are tagged as Go-Lab and Next-Lab project members (as they are assumed to be more knowledgeable about IBL than other teachers that may have entered the platform through trainings and other means). The rest of the users registered in the authoring platform are considered potential *teachers*. Also, an ILS has been considered *implemented* in the classroom (i.e., adopted) when more than 10 “student-mode users” are registered to it (this empirical rule of thumb stems from conversations with teachers using the platform, as their classes have on average around 20 students, and teachers often report their students using the platform in groups of 2 or more people).

Regarding data analysis, basic descriptive and exploratory statistics have been used. First, indicators have been drawn for each ILS as markers for the presence of different KAM social practices (see the following section). Then, counts of ILSs (and proportions of their being implemented in classrooms) have been calculated, for different levels of maturation and social involvement. Furthermore, Chi-squared tests of independence and logistic regression modelling have been performed to ascertain the potential association between these KAM-related practice indicators and ILS implementation (and their relative strength).

## Study

### Knowledge appropriation and adoption in Go-Lab

In order to better understand how KAM practices are materialized in Go-Lab, this section describes how teachers deal with knowledge maturation, scaffolding and appropriation, reflecting on where these practices take place (inside or outside of the technical ecosystem), as well as whether and how they can be monitored. Figure 1 provides a graphical overview of this, and more concrete metrics are listed in Table 1.

**Knowledge maturation practices.** With the help of the technical ecosystem, teachers can *appropriate* existing IBL templates when creating their ILSs from scratch. Teachers can invite other peers into their ILSs either for *sharing* or for *co-creation* purposes. Once an ILS is ready, teachers can publish it to make it available for other teachers. This process also involves an expert review that provides feedback via email about how to *formalize* the ILS. *Standardization* happens when ILSs are widely uptaken by other teachers at the national or international level to address specific parts of the curriculum.

**Knowledge scaffolding practices.** Teachers may *request help* from experts using the help desk available in the ecosystem, where a set of experts is available to answer doubts and provide *guidance*. Additionally, teachers often address their trainers or national coordinators face-to-face, via email, video conference, social media, or inviting them into their ILSs, so that they can have a look, comment or even edit the ILS. In addition, there are many cases where experts accompany teachers while implementing ILSs for the first time. Then, to encourage teachers to solve their own problems and be autonomous, expert support *fades* progressively reducing the face-to-face support and then interventions in the ILSs.

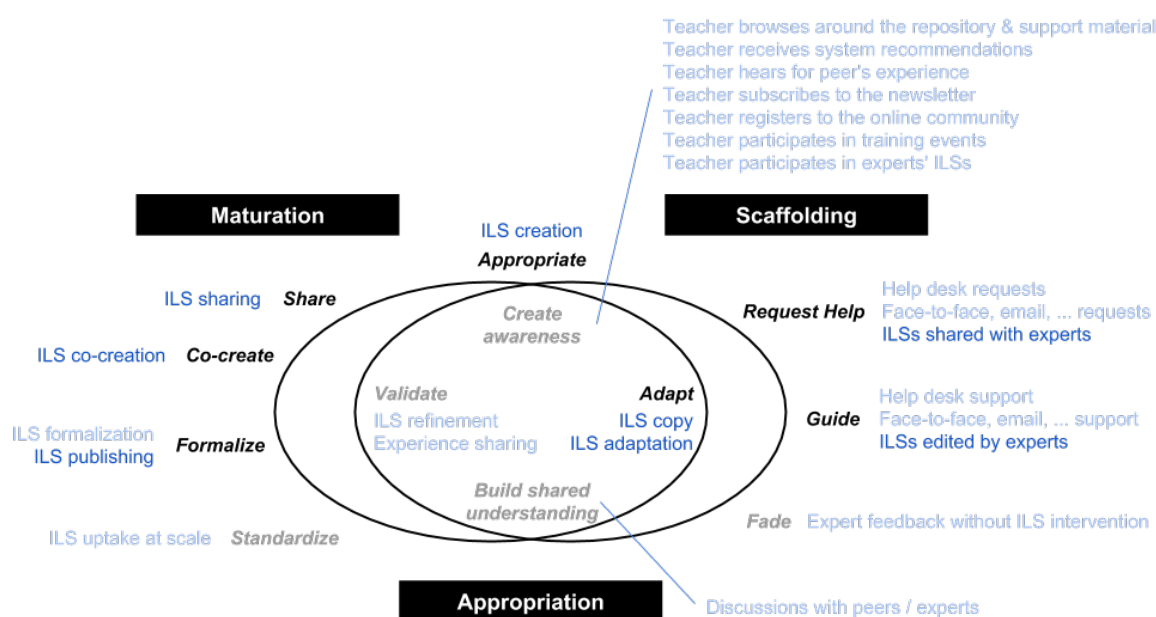
**Knowledge appropriation practices.** Teacher *awareness* comes from four different sources: the teachers themselves, searching and browsing through the repository and the support material; the technical ecosystem, which provides with recommendations about relevant apps, labs, resources, etc.); peers, who share their knowledge or experience using communication channels outside the ecosystem (e.g., face-to-face or in the social media); and, experts, who keep teachers updated about the new apps, labs, ILSs, functionalities and events through the newsletter, the online community and, especially, through training events. Once aware of the

available support, teachers can *adapt* existing ILSs to their own educational contexts. Thus, we consider the ILS reuse as a general proxy for the fact that knowledge appropriation has taken place.

Although the ecosystem supports communication between users, most of the discussions between peers to reach a common *understanding* about the learning context and the ILS take place outside of the platform. The review process before the ILS publication triggers a conversation between experts and teachers that not only leads to ILS refinements, but also improves the experts' *understanding* about teacher needs, usability issues, misconceptions and challenges for ILS adoption in the real classroom.

Regarding the ILS *validation*, this practice usually takes place during the training events (out of the technological ecosystem) where practitioners are invited to share their experiences implementing ILSs in the classroom. Also, through iterative refinements due to the reuse, the validation of the content of an ILS increases copies (and copies-of-copies) of it are used and implemented.

**Innovation adoption in Go-Lab.** While there are different degrees of adoption (e.g., teachers can use IBL practices, labs, or apps independently), in this paper we focus on the potential usage of ILSs in the classroom since they represent the overall Go-Lab proposal: the IBL pedagogical approach combined with apps, labs, and multimedia resources that contribute to scaffold the students in the learning process.



**Figure 1.** Knowledge Appropriation Model (in italics) and the corresponding practices in Go-Lab. Those practices that cannot be mapped with monitorable metrics in Graasp appear in lighter color.

While these knowledge appropriation practices involve both face-to-face and computer-mediated interventions, in this paper we focus on those supported (and traceable) through Go-Lab's technological ecosystem. For instance, since the Go-Lab repository does not require logging in, no evidence can be obtained about the "create awareness" actions carried out by specific teachers there. Therefore, only the practices mediated by the authoring platform, i.e., Graasp, will be taken into consideration in the analysis.

## Results and discussion

### Hypothesis 1: The ratio of adoption increases across knowledge maturation categories

First, we explored whether the adoption of an ILS was dependent on knowledge maturation practices. Specifically, we checked whether the fact that an ILS was assumed to be in higher stage of maturity had an effect on its eventual adoption in the classroom. In order to test the first hypothesis, *the ratio of adopted ILSs would increase across the categories of maturation* (individual, shared, co-created, formalized), each ILS created by a teacher was assigned to one exclusive knowledge maturation category according to the indicators scheme shown in Table 1. Then, for each maturation category, the rate of adoption was calculated.

Table 1: Monitorable knowledge maturation levels versus adoption (ILS implementations) in Go-Lab

| Maturation level | Indicator  | Adopted | Not adopted | % Adopt. |
|------------------|--|---------|-------------|----------|
| Individual       | ILS was created and neither shared nor published       | 651     | 21631       | 2.92%    |
| Shared           | ILS shared with peers but not co-created nor published | 85      | 560         | 13.18%   |
| Co-created       | ILSs co-created with peers, but not published          | 293     | 745         | 28.23%   |
| Formalized       | ILS was published for the community                    | 188     | 332         | 36.15%   |

Our results show that knowledge maturation is strongly related to later adoption of the ILS in the classroom. While it is true that the accessibility is restricted, i.e., the individual spaces are not widely accessible in Graasp, we still find it remarkable that collaboration seems to be strongly related to adoption. Especially outstanding is the difference between sharing and co-editing. In this case, the accessibility would be typically the same, i.e., the same amount of users would have access, but the adoption more than doubled between sharing and co-editing. Aside from these descriptive results, a Chi-squared test of independence also indicates such strong association between both traits of an ILS ( $\chi^2=2549.2$ ,  $p<0.001$ ).

#### Hypothesis 2: The ratio of adoption increases when knowledge is appropriated

Since the KAM assumes knowledge appropriation to be a prerequisite for knowledge maturation and eventually adoption, we hence hypothesized that knowledge appropriation (in this case the reuse of an ILS) should be one of the determinants of maturation. In the current setup of the system, it was not feasible to gather data about individual practices of appropriation (creating awareness, shared understanding, adapting and validating), and thus for this analysis we considered whether a particular ILS was reused (copied and adapted from another ILS), as a proxy for whether knowledge appropriation had taken place. Then, as shown in Table 2, we compared the rates of adoption for reused versus non-reused ILSs.

We found that knowledge appropriation in Go-Lab also is strongly associated with later adoption. In this case, we assumed that the reuse of an ILS would be indicative of several appropriation practices: to reuse existing material, there should be first awareness of its existence; then, to implement the ILS in the classroom, it is necessary to refine and adapt it to the contextual needs, contributing to build sharing understanding; and the adoption shows not only the interest of the community but also validates the feasibility to apply it in real scenarios. A Chi-square test of independence also rejects strongly the null hypothesis (i.e., appropriation and adoption being independent):  $\chi^2=2035.9$ ,  $p<0.001$ .

Table 2: Monitorable knowledge appropriation levels versus adoption (ILS implementations) in Go-Lab

| Appropriation level | Indicator                              | Adopted | Not adopted | % Adopted |
|---------------------|--|---------|-------------|-----------|
| Not reused          | ILSs of which no copies have been made | 868     | 22642       | 3.69%     |
| Reused              | ILSs of which copies have been made    | 349     | 626         | 35.79%    |

#### Relative associations between knowledge maturation and appropriation

In order to perform a first exploration of the relative strength of the two variables targeted by our hypotheses (i.e., knowledge maturation and appropriation practices), we built a logistic regression model that predicts whether an ILS will be adopted or not, as a function of its levels of maturation (from individual to formalized) and appropriation (i.e., whether it is a re-use of a previous one). The results are summarized in Table 3. The results seem to indicate that both maturation and appropriation have a significant positive effect (as indicated by the analysis of the hypotheses above): each level of maturation or appropriation increases the odds of an ILS being implemented by a factor of 3.7 and 3, respectively. Moreover, the fact that the quadratic coefficient of the knowledge maturation is negative (i.e., the positive association decreases in higher levels of maturation) may point to the fact that maturation is especially important in the early phases of the process, where the artifacts still need a lot of collaborative effort in order to be understood and usable for others.

Table 3: Coefficients of a logistic regression model of adoption as function of appropriation and maturation

| Coefficients                       | Estimate | Std. Error | z value | p-value |
|------------------------------------|----------|------------|---------|---------|
| Intercept (Individual, Not reused) | -1.463   | 0.045      | -32.23  | <0.0001 |
| Maturation (linear)                | 1.298    | 0.097      | 13.32   | <0.0001 |
| Maturation (quadratic)             | -0.978   | 0.089      | -10.92  | <0.0001 |
| Appropriation (linear)             | 1.097    | 0.076      | 14.36   | <0.0001 |

## Conclusion and future work

While social processes have been considered crucial in teachers' ongoing professional learning, their implicit and informal nature make their study challenging. In this paper, the Knowledge Appropriation Model (KAM) helped us to define traceable social practices that could then be related with the adoption of IBL practices in the classroom through Go-Lab. This focus on social practices at work makes our study closely related to the computer-supported collaborative work domain. However, we should not forget that the main focus is on the teachers learning to integrate IBL in their practice, and how informal (computer-supported) collaborations around shared artifacts (in our case ILS) facilitate this learning. Then, the usage of these artifacts in the classroom should be looked at as one of the few reliable markers to assess whether teachers have learned to integrate such pedagogy in their practice.

Based on the ILSs created between November 2015 and November 2018, out of 24,485 ILSs created by teachers, 1270 (4.99%) were potentially implemented. The descriptive and exploratory analyses of the dataset support our initial hypotheses about how collaborative professional learning practices (as defined by KAM) relate to teacher adoption of IBL pedagogy in Go-Lab: the higher the maturation and the appropriation, the higher the adoption – as a marker of teachers learning the pedagogy. This positive association between knowledge maturation, appropriation and IBL adoption, could help to enrich existing strategies for teacher professional development in IBL (Maaß & Doorman, 2013). For example, mentoring and collaborative practices among teachers could be promoted in trainings, and be facilitated even further in IBL digital platforms.

While these findings are encouraging, other factors may well be playing a role in the adoption (e.g., participation in trainings, previous IBL teacher experience, experts' interventions, etc.). Indeed, the present study focuses on social processes mediated by an on-line platform, whereas additional collaboration may have taken place through other face-to-face or digital channels. Thus, further research (e.g., in the form of follow-up qualitative studies) is needed to examine the assumptions and conclusions of our present analysis, and the influence that other factors may have had on the adoption of Go-Lab.

It should be also noted that the filter applied to detect ILS adoption was purely superficial (based on the average number of students expected per classroom implementation) and did not take into consideration the qualitative characteristics of the ILS. Therefore, future studies should analyze more thoroughly the work done by the students as well as the ILS structure and content. This includes not only content/structure analyses of the ILSs, but also time-based models of the knowledge maturation and appropriation processes (e.g., using process mining techniques on the different stakeholders' actions in the platform).

Another feature of this initial work towards disentangling the different social processes that play a role in the adoption of IBL, is our present focus on the professional learning artifacts (i.e., the ILSs), rather than focusing on the teachers themselves as the unit of analysis. In future work we expect to find models that help us understand not only artifacts but also skill adoption among practitioners, so that we can identify and promote efficient learning paths for teachers.

While the overall results described in this paper (i.e., the fact that more mature artifacts/lessons are implemented more often in the classroom) may seem somewhat self-evident, our main contribution is rather the operationalization of this hypothesis in terms of concrete, traceable metrics in a digital platform. This is especially valuable in the realm of (collaborative) professional learning, where empirical evidence of such learning and surrounding collaborative processes tends to be scarce. The first steps presented here open thus the door to further investigations of how collaboration around shared, “implementable in practice” artifacts mediate informal professional learning.

## Endnotes

- (1) Knowledge Appropriation Model: <http://results.learning-layers.eu/scenarios/knowledge-appropriation>
- (2) Go-Lab initiative: [www.golabz.eu](http://www.golabz.eu)
- (3) Graasp: [graasp.eu](http://graasp.eu)

## References

- Billett, S., & Choy, S. (2013). Learning through work: emerging perspectives and new challenges. *Journal of Workplace Learning*, 25(4), 264–276.
- Blumenfeld, P. C., Krajcik, J. S., Marx, R. W., & Soloway, E. (1994). Lessons Learned: How Collaboration Helped Middle Grade Science Teachers Learn Project-Based Instruction. *The Elementary School Journal*, 94(5), 539–551.
- de Jong, T., Linn, M. C., & Zacharia, Z. C. (2013). Physical and virtual laboratories in science and engineering education. *Science*, 340(6130), 305–308.
- Fairweather, J. (2008). *Linking evidence and promising practices in science, technology, engineering, and mathematics (STEM) undergraduate education*. The National Academies National Research Council Board of Science Education. Retrieved from
- Henderson, C., & Dancy, M. H. (2011). *Increasing the impact and diffusion of STEM education innovations*. Presented at the White Paper commissioned for the Characterizing the Impact and Diffusion of Engineering Education Innovations Forum. Retrieved from <https://www.nae.edu/File.aspx?id=36304>
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist*, 41(2), 75–86.
- Lee, V. E., Dedrick, R. F., & Smith, J. B. (1991). The Effect of the Social Organization of Schools on Teachers' Efficacy and Satisfaction. *Sociology of Education*, 64(3), 190–208.
- Maaß, K., & Doorman, M. (2013). A model for a widespread implementation of inquiry-based learning. *ZDM: The International Journal on Mathematics Education*, 45(6), 887–899.
- Maier, R., & Schmidt, A. (2015). Explaining organizational knowledge creation with a knowledge maturing model. *Knowledge Management Research & Practice*, 13(4), 361–381.
- Nonaka, I., von Krogh, G., & Voelpel, S. (2006). Organizational Knowledge Creation Theory: Evolutionary Paths and Future Advances. *Organization Studies*, 27(8), 1179–1208.
- OECD. (2014). *TALIS A Teachers' Guide to TALIS 2013 Teaching and Learning International Survey: Teaching and Learning International Survey*. OECD Publishing.
- Paavola, S., & Hakkarainen, K. (2014). Triological Approach for Knowledge Creation. In S. C. Tan, H. J. So, & J. Yeo (Eds.), *Knowledge Creation in Education* (pp. 53–73). Singapore: Springer Singapore.
- Rodríguez-Triana, M. J., Govaerts, S., Halimi, W., Holzer, A., Salzmann, C., Vozniuk, A., ... Gillet, D. (2014). Rich open educational resources for personal and inquiry learning: Agile creation, sharing and reuse in educational social media platforms. In *2014 International Conference on Web and Open Access to Learning (ICWOAL)* (pp. 1–6). IEEE.
- Schenke, W., van Driel, J. H., Geijsel, F. P., Sligte, H. W., & Volman, M. L. L. (2016). Characterizing cross-professional collaboration in research and development projects in secondary education. *Teachers and Teaching*, 22(5), 553–569.
- Seymour, E., Hunter, A.-B., Laursen, S. L., & DeAntoni, T. (2004). Establishing the benefits of research experiences for undergraduates in the sciences: First findings from a three-year study. *Science Education*, 88(4), 493–534.
- Slotta, J. D., Tissenbaum, M., & Lui, M. (2013). Orchestrating of Complex Inquiry: Three Roles for Learning Analytics in a Smart Classroom Infrastructure. In *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (pp. 270–274). Leuven, Belgium.
- Teo, T. (2009). Modelling technology acceptance in education: A study of pre-service teachers. *Computers & Education*, 52(2), 302–312.
- Vuorikari, R., Berlanga, A., Cachia, R., Cao, Y., Fetter, S., Gilleran, A., ... Others. (2011). ICT-based school collaboration, teachers' networks and their opportunities for teachers' professional development—a case study on eTwinning. In *International Conference on Web-Based Learning* (pp. 112–121). Springer.
- Webb, M., & Cox, M. (2004). A review of pedagogy related to information and communications technology. *Technology, Pedagogy and Education*, 13(3), 235–286.
- Wenger, E. (1999). *Communities of Practice: Learning, Meaning, and Identity*. Cambridge University Press.

## Acknowledgments

This research has been partially funded by the European Union in the context of Go-Lab, Next-Lab, CEITER (FP7 grant agreement no. 317601, and Horizon 2020 Research and Innovation Programme, grant agreements no. 731685 and 669074).

# A Scaled Analysis of How Minecraft Gamers Leverage YouTube Comment Boxes to Participate and Collaborate

Justice T. Walker, Stefan Slater, and Yasmin B. Kafai  
justicew@upenn.edu, slater.research@gmail.com, kafai@upenn.edu  
University of Pennsylvania Graduate School of Education

**Abstract:** Minecraft—a commercially successful and popular sandbox game—is increasingly used in formal educational spaces. While there is significant research on how online gaming can support positive outcomes when used as a pedagogical tool, little research has investigated how gamers collaborate to learn in informal spaces. We conducted an analysis of a curated subset of more than 540,000 comments of Youtube videos to explore how users participate and practice in discussions about Minecraft where comment boxes are the primary method used to communicate about video content. Our findings reveal that gamers use online comment sections for multiple reasons, from participating in shared social experiences unique to Youtube as a community to giving myriad forms of feedback to learn and further content development. In our discussion, we address the ways other online game platforms can better support learners seeking out information and collaborating in ways that are constructive and inclusive.

**Keywords:** Gaming, Metagaming, Feedback, Participation, Minecraft

## Introduction

Much research has focused on learning and collaboration while playing (Connolly et al., 2012; Clark et al., 2016) and making digital games (De Freitas, 2006; Basawapatna et al., 2010), which has collectively pointed to the influence games have in supporting positive outcomes in academic content areas (e.g., language arts, computer science, mathematics) and skills (e.g., literacy, problem solving and argumentation). These outcomes have been especially promising for education efforts designed to support broader access to science fields where social and intellectual participation disparities exist. Alongside these efforts is research that attends to learning and collaboration in massive multiplayer games where game elements and social interactions are complex and more difficult to ascertain (Paraskeva et al., 2010). Examples include commercially successful games like the role playing-based World of Warcraft (Clark et al., 2009), the simulation-based SimCity (Tanes and Cemalcilar, 2010), and the open narrative-based Minecraft (Nebel et al., 2016) where gamers can interact on massive server networks that are globally distributed.

With more than 100 million registered users, Minecraft is an example of a game where users are able to collaborate to make and play games (Makuch, 2014). In Minecraft, players use tools and materials to build complex structures that can be automatized or programmed, especially through the use of third-party game modifications (mods). YouTube has become a popular metagaming environment for Minecraft by providing an online open access video platform where content creators have posted more than 70 million Minecraft tutorials videos publicly. In fact, Minecraft-related searches on Youtube represent the second most commonly searched term on the video platform (Thompson, 2016). Prior research of this environment examined a small number of Minecraft player Youtube video producers to understand the nature of user interactions (Niemeyer and Gerber, 2015). Findings suggested that this online video platformed represented—in broad terms—a “collaborative learning community” that supported constructive engagement among players. While these findings offer important justifications for further research on how collaboration takes place in this metagaming space, few research (Niemeyer and Gerber, 2015; Anderson et al., 2017) studies have focused on learning outside of gaming environments and in digital spaces that develop around games—the so-called metagaming spaces (Gee, 2008). Gee and others have argued that more learning happens in these contexts and communities than in traditional learning spaces. However, research of these spaces is difficult: metagaming environments generally operate on private web servers (e.g., personal networks or in-game chat systems) or are distributed across physical environments (e.g., conferences, camps, festivals, etc) that represent only a small fraction of the total number of community members that participate in a given game. As a result, it has been difficult to research these contexts at scale. In this study, we use natural language processing (NLP) approaches to characterize and make sense of the complex ways Minecraft gamers participate and provide feedback when using video tutorials and comment sections. We ask: (1) what modes of participation are used by Minecraft players in the comments section of Youtube? (2) what forms of feedback do Minecraft users provide when using Youtube? and (3) what

are the affordances and constraints of using online video platforms to participate in feedback? In the discussion, we address the ways online game platforms can better support accessible and equitable learning and collaboration in game environments that are self-directed as well as directions for future research.

## Background

### Collaboration in gaming

In learning, collaboration has been described as moments when learners and/or experts come together to share ideas, feedback, and attention toward a shared understanding or goal (Davis et al., 2018). The relationship between collaboration and learning has been conceptualized using constructivist perspectives that describe learning as a social interaction where so-called novices and experts come together to learn from one another. In this theoretical trajectory, digital games can be understood as communities of practice as players of different mastery levels convene around a common interest (Lave and Wenger, 1991). It is within these communities where collaborations guide interactions and support learning. There are a number of collaborative arrangements in digital game environments that positively support these outcomes and that have been shown to be instrumental in important cognitive processes including reflection.

One collaborative arrangement that has been shown to be especially effective involves situations where learners provide each other feedback (Butler and Winne, 1995; Plass et al., 2015). This has been effective in traditional learning research in literature, mathematics, and computer science. Feedback has also been examined in digital game platforms. Research in this area has shown that collaborative feedback has supported (civic) engagement, knowledge building, and skill building and across a range of subject areas. While most studies examine collaborative feedback while playing games, this research explores a new genre of game platform that combines game play and construction. In this form, gamers not only engage in play, but they also have opportunities to build and modify game environments, characters and objectives.

Minecraft is an example of this type of game and is the frame within which this research examines collaboration and feedback. Platforms like Minecraft have previously been described as among the most promising game-based approaches to support learning (Gee, 2008; Plass, 2015). However, education research that explores the ways Minecraft can support learning is scant (Nebel et al., 2015). Existing research has described ways the ways Minecraft can be utilized in K-12 learning environments to teach mathematics concepts (Short, 2012), in-game collaboration tactics (Marklund et al., 2013) and collaborative design (Cipollone et al., 2014; French et al., 2016). Research has also assessed collaboration differences in underrepresented minority groups (Ames & Burrell, 2017), and small-scale analysis of collaboration techniques for learning (Wernholm & Vigmo, 2015; Davis et al., 2018) among young players.

### Collaboration in metagaming

While these studies provide important insights into the nature of collaborations among diverse players, much less is known about how gamers collaborate in spaces outside the actual game platform—Minecraft’s metagame environments. Because these environments play a significant role in supporting learning outside of Minecraft (Bebington, 2014), understanding the ways gamers collaborate in these spaces provides perspectives that are not always apparent in the game environment. Although research has shown that feedback benefits collaborative learning arrangements, adverse outcomes have also been identified. Some of these outcomes include inequitable participation where only a few learners benefit from learning outcomes associated with feedback-driven collaboration (Ames and Burrell, 2017). Other adverse outcomes occur when feedback is not constructive or is not used to drive reflection. Because of challenges related to access and scale, little research investigates the myriad ways in which learners engage in feedback in digital games or their associated metagaming environments.

Videos on Youtube can provide an important insight into the nature of collaboration in the metagaming space. Minecraft players post millions of videos on YouTube showcasing their play experience or explaining particular concepts. Minecraft related topics on Youtube include instructions for making and applying ‘skins’ (changes to the physical appearance of a Minecraft avatar), strategies for mining and obtaining resources, redstone networks based on computational circuits and logic, and ‘modding’ (third-party code that adds additional game functionality).

Previous research that examined thousands of comments in the online youth programming community Scratch (Fields et al., 2015) found that players were able to provide constructive feedback. However, within the Scratch online community, the commenting feature is integrated within the environment and thus not a traditional metagaming location. Youtube videos are a traditional metagaming location, but the traffic served by the website is multiple orders of magnitude larger. To overcome this challenge, we use a bag-of-words



approach—specifically keyword spotting (Cambria and White, 2014)—to curate, collect and analyze moments of user collaboration in order to provide a scaled insight into collaborations that occur in these comment spaces, the collaboration affordances and constraints that exist using online video platforms, and ways to support constructive collaborations in similar online video-based meta game environments.

## Methods

Data were collected in September 2018, from the comment sections of a sample of videos posted to Youtube. In order to generate our sample of videos, we first developed a list of search terms related to different actions and behaviors within Minecraft that users might want to engage with. We used Youtube's public application program interface (API) to write a Python script that returned the top 50 video results for each of our search terms, detailed in the next section. We used the video identification (IDs) obtained from this process to collect every comment left on each of the sampled videos. This process yielded more than 540,000 unique comments. From this sample we performed qualitative coding using an inductive approach (Ravitch and Carl, 2015) in order to identify a subset of terms relevant to learning and social engagement. This yielded a sample of more than 27,000 unique comments. Of this sample, we randomly selected a subset of about 100 and developed a ground-up codebook to describe how Youtube users participate and collaborate with one another when using the comments section to learn. We then applied these codes to a random set of 518 unique comments drawn from the 27,000 comment sample. In the following section, we detail our specific video sampling criteria and subsampling techniques.

## Video selection

We searched Minecraft-related videos based on nine search terms encompassing a number of popular approaches to playing. For each search, the top 50 video results according to Youtube's relevance algorithm were selected. From these videos, every comment was recorded along with the author, the date posted, the content of the comment, and the number of likes/replies the comment received. To collect comments related to 'skins,' we searched for videos using the search terms 'minecraft how to skins' and 'minecraft skins tutorial.' For mining, we used search terms 'minecraft how to mine' and 'minecraft mining tutorial'. For 'Redstone,' we used the search terms 'minecraft how to redstone' and 'minecraft redstone tutorial.' Finally, we used the search terms 'minecraft best mods', 'minecraft how to mod', and 'minecraft mod tutorial' to search for videos related to the modding process in Minecraft.

## Subsampling the dataset

Including several videos collected during our code testing, our dataset included 433 unique videos and 546,034 unique comments. Due to the size of this dataset, we generated several smaller datasets to make qualitative coding approaches feasible. We performed this dataset reduction in two steps. We first selected all comments from the dataset that were either liked or replied to at least once. These criteria allowed us to select for comments that showed some level of social engagement. This yielded two datasets—52,775 'liked' comments and 37,930 replied comments, respectively. In the first dataset, 21,441 comments were liked more than once, and the maximum number of likes received by any comment was 9,228. In the reply dataset, 18,334 comments were replied to more than once, and the maximum number of replies received by any comment was 7,460.

We used these two datasets to perform preliminary inductive coding, in order to identify how Youtube users engage with Minecraft, and how collaboration and learning within the comments might be operationalized. From this preliminary coding, we identified a set of 11 key words and phrases that appeared to be associated with a deeper level of engagement in the content. These keywords included 'figure out' (433 comments), 'turn out'/'turns out' (55), 'I found'/'I have found' (1,849), 'because' (8,440), 'I think' (4,381), 'I can' (6,791), 'how do' (4,371), 'how can' (656), and 'learn to'/'learn how'/'learn about' (176).

Finally, we performed both inductive and deductive coding on each of these keyword datasets. For datasets containing more than ~100 comments, we selected random samples of 100 comments from within them to code. We used an inductive coding approach to identify both common themes of discussion on Youtube, as well as social norms and platform-specific memes that users engage in. Deductive coding was used to identify the degree to which a comment was emblematic of involving feedback or social participation.

## Findings

We report finding about the forms of participation Minecraft gamers use when using the comments section on Youtube tutorial videos. We also report on the various ways Minecraft gamers use feedback to collaborate and participate. Unless otherwise stated, these codes were not mutually exclusive and could co-occur.

## Finding 1: Minecraft gamers mostly: use comments to participate, reference in-game activities, and engage Youtube video producers

Of 518 Youtube comments assessed, 346 involved some form of social participation with other people. This could involve either sharing personal experiences from playing Minecraft, or it could involve conversations with video authors or other users on Youtube.

101 of these participation comments were coded as involving *ways that Minecraft gamers participated*. These results are summarized in figure 1a. Unsurprisingly, in-game references (88%) represented the largest form of participation. These were moments when gamers made reference to experiences either playing Minecraft or engaging with Minecraft. For example, one commenter explained, “So aswome because I don’t mine at night any more.but when I go in the day, it turns out night when I get out an have to sprint to get to my house.” Here the user is using the comment box to describe their strategy during play after viewing a related tutorial. References to personal experiences (12%) also occurred and represented instances when gamers mentioned personal life experiences that were not directly related to Minecraft. For example, a video reminded a commenter of sleepovers with a friend, “I first heard this 4 years ago when my friend came over and introduced me to minecraft. 4 years later and I have found the song. OMG so much nostalgia. Plus also nostalgic because it reminds me of all the sleepovers I had with the friend.” In this example, it can be seen that personal stories often overlapped with nostalgia (11%), which was represented in moments when gamers made reference to or recollected a time in the past. Finally, the smallest set of comments represented instances when gamers used comment boxes to reference to other games (2%).

201 comments were addressed to a particular audience, and explicitly social in nature. The majority of these comments were directed toward Youtube video producers (53%). For instance, as one commenter used the comments to solicit help from the video author directly, saying “I really wanted a mod but it turns out ur tutorial doesn’t work for me when I Vick ur link it takes me to a different forge pls answer me.” Other commenters (13%) and Minecraft game players broadly (2%) were also focal points of engagement in comment boxes as commenters would ask others to describe their own experiences or refer to other gamers on Youtube about Minecraft experiences. Finally, we coded for unclear or ambiguous participatory comments, instances where it was not possible to determine to whom the engagement was directed. For instance, the comment “you know what do you have a server ? because there is a cool game its called egg wars:D” could be directed either at a video author or another commenter.

96 comments referenced Youtube as a platform and its various design affordances. Comment codes (85%) included references to using Youtube’s comments section as a site of shared participation, such as requests to “start a comment chain” around a particular idea or behavior in the video. This represented the largest percentage of participation forms gamers used on Youtube. References to comment or video “likes” (7%) are instances where the commenter refers to the number of likes a video has or encourages other users to like/dislike a particular video. Technical references (6%) were also observed and were instances when gamers referenced a technical aspect of Youtube such as lag on a video or a missed notification. References to subscribing (5%) to a video were also observed and involved mention of the subscribe feature of a Youtube video channel. These comments were often solicited by the author (e.g., “SUBSCRIBE!!!”) in order to grow a viewer audience base. Notifications (1%) were infrequently observed and represented instances when a gamer refers to being notified (or not) about video content on Youtube. Similar to subscribe references, these typically represented moments when commenters signaled wanting to be a part of or being a part of a viewer audience and stay up to date on new content.

Finally, 80 comments included digital artifacts of participation beyond simply leaving a comment: poems (68%), hashtags (14%), stories (10%), emojis (4%), memes (4%), and links to external resources (3%). These results are summarized in figure 1b. These represent instances when gamers used atypical representations to participate, such as collaboratively posting lyrics to a song parody, or telling elaborate and embellished stories about a particular behavior.

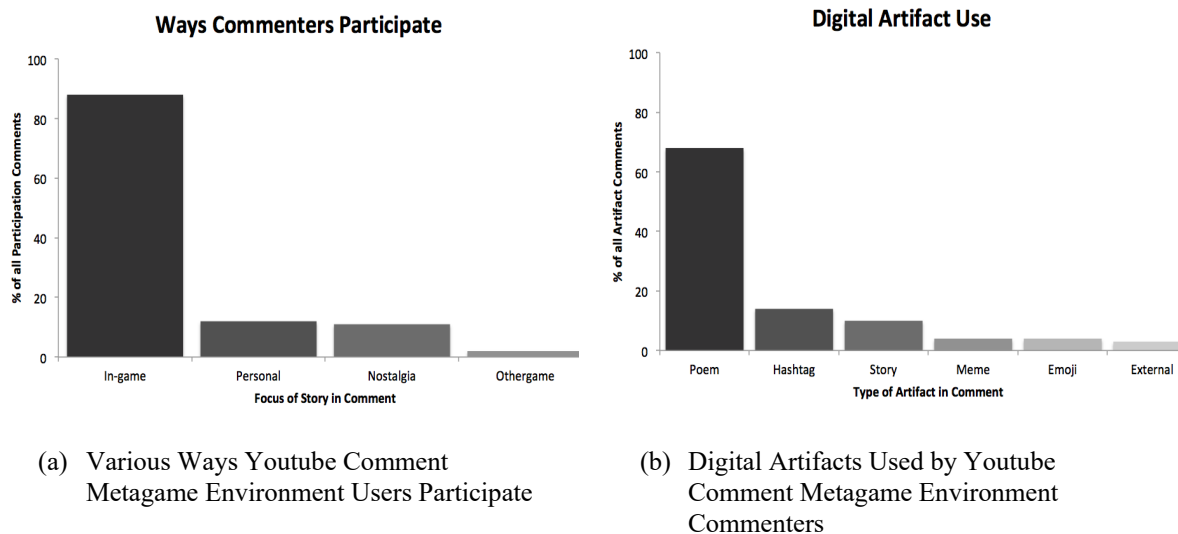


Figure 1. Participation Forms (a) and Digital Artifacts (b) Used.

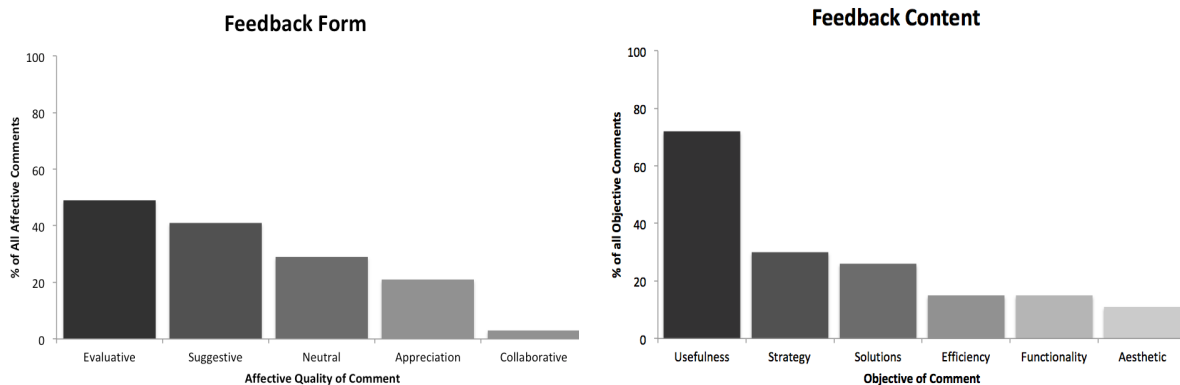
## Finding 2: Minecraft gamers mostly use affective and neutral feedback and primarily to assess Youtube video usefulness and functionality

261 of the 518 coded comments involved some form of feedback, with users sharing their thoughts and opinions on video content, video authors, or other commenters, among other topics

Of this sample, 134 involved comments that referenced specific in-game game features including ‘mods’ (41%), redstone (37%), mining (19%), building (18%), general play (16%), and ‘skins’ (7%).

Also, from this sample, 168 comments reflected instances when gamers used the Youtube comment section to provide either affective or social feedback. Feedback that was evaluative (49%) included judgements that included affective features like sarcasm, insults, compliments, or encouragement. Feedback that was suggestive (41%) included moments when commenters suggested ideas for future Minecraft-related tasks, strategies, and/or video content. An illustrative example of this occurred when a commenter remarked, “can you do a customstom command seiries were you learn to make a command block machine and make minecraft a better place to play in :).” Here the commenter is suggesting through request that the Youtube video producer make a specific series of tutorial content. Neutral feedback (29%) reflected moments when commenters provided feedback that did not have an affective quality. This was illustrated when a commenter remarked about the audio being played during a tutorial “shoulve been learn to fight at night,” which is a critique about the audio that best represents the tutorial and yet does not explicitly connote any affective judgement. Feedback that was appreciative (21%) reflected moments when commenters asserted thanks because of something related to video tutorial content or the video producer. This type of feedback was evident when a commenter noted, “Thanks for this series seth it is really clear and instructive and i’m really looking up to learn to code turtles makes FTB way easier.” The commenter’s feedback also goes further to explain the reasons why they appreciated the tutorial. Feedback that was collaborative (3%) in nature demonstrated moments when video tutorials prompted a commenter to solicit video producers or other commenters to collaborate on a common Minecraft related endeavor. Findings are summarized in figure 2a.

Content focused feedback occurred in 112 feedback-related comments. Of these, the majority emphasized video tutorial usefulness (72%) in helping viewers achieve some in-game outcome as illustrated when a commenter remarked, “Hey Mumbo I love your vids and watch them every day! Your redstone tutorials have helped me sooo much and I have actually have been able to build some complicated redstone contraptions because of you! [emojis]” In this case the commenter is underscoring the idea that the video producer’s tutorial helped the content viewer achieve a goal - in this case learning to use redstone in complicated ways. These observations are summarized in figure 2b. In addition, 172 involved comments that reflected feedback that was directed toward: content (88%), Youtube producers (11%) and technical video features (3%).



(a) Forms of Feedback Provided on Youtube Metagame Environment Comment Boxes

(b) Feedback Content Provided on Youtube Metagame Environment Comment Boxes

Figure 2. Feedback Focus (a) and Forms (b).

There were also instances when commenters provided feedback about the strategy (30%), solution (26%), and efficiency (15%) of proposed game play. An illustrative example of these occurred when one commenter asserted, “Dude searching all this stuff is slow as hell and less reliable, sure mining sometimes turns out slow but trust me after like 40 minutes I have stacks of materiel and sometimes enough for Diamond Armour.” Here, the commenter is providing feedback about the efficiency of a mining strategy and the benefits of using the approach in reference. Feedback related to functionality (15%) reflected moments when video commenters made reference to the function of an approach such as the ability to produce a functional redstone-based product. Aesthetics (11%) related feedback were moments when commenters referred to the aesthetic quality of a video or Minecraft artifact such as—for example—the sound of a song or look of a building as exemplified when one commenter noted “I like the song, but i think the vid would be better with the default textures....” The commenter is pointing out the mod used to alter the game environment would have been visually better if the the game’s default settings were maintained.

## Discussion and conclusion

Here we advance a mixed-methodological approach that uses NLP and qualitative methods to overcome challenges associated with analyzing qualitative data from metagaming platforms at scale. By doing so, we’re able to assess the myriad ways Minecraft users participate and collaborate in tutorial-based video content on Youtube. Our analyses show several trends involving the types of feedback and collaboration that occur on this platform.

First, we observed that Minecraft gamers use Youtube to participate in multiple forms of discourse. While comment section discourse represented the most frequent form of engagement in our sample, we also observed song remixing and parodying. This suggests that online platforms support unique cultural norms and practices. We also observed moments where Youtube supported just-in-time, on-demand access to quasi-experts—video content producers that have specialized domain knowledge of a set of Minecraft related features (e.g., redstone, mods, skins, etc.). This underpins important implications for learning at scale in environments where access to domain experts is limited. Here, we observe that metagame communities provide an important space for learners to not only engage with experts in order to achieve learner-centered goals in the game environment, but also to provide feedback on content which these experts should cover.

We also observed that Minecraft gamers frequently use forms of feedback to assess Youtube content. Notably, feedback about Youtube videos primarily focused on game-related content, suggesting that this metagame environment is an important space for (Minecraft) learning—harkening back to Gee’s characterization of the ways these spaces support meaningful engagement. Furthermore, commenters most often provided feedback that was evaluative and suggestive. This underpins the prevalence of affective social engagement on Youtube. It suggests that Youtube acts as a place for learners to advance not only their own learning repertoires, but also that of others. Learners very often made unprompted suggestions (both to video authors and other commenters) to advance tutorial content.

Importantly, this research provides important first attempts to elucidate insights into the myriad ways gamers use a metagaming environment experience to participate and productively collaborate while learning.

While Youtube affords opportunities for diverse participation and engagement, important constraints exist. These constraints include the ability for users to participate in discourse beyond those afforded by Youtube's technical features. We observed, for instance, users participating and collaborating in ways that were marginal compared to prominent comment-based and evaluative forms. These instances represent an important need to create space for participants to engage in more broad forms of engagement. In addition, our research uncovers an important risk that exists when social engagement takes on a predominantly affective orientation. This risk could create adverse outcomes if, for instance, dominant voices or practices unintentionally place groups at risk of marginalization in this space, an outcome that could exacerbate education disparities. Nevertheless, we report on methodological approaches and perspectives that support research on metagame environments at scale.

In conclusion, we argue that metagaming environments are spaces where important learning processes take place. The affordances of these environments can be used by educators as well as learning systems designers in order to produce more effective and fulfilling collaborative learning practices.

## References

- Ames, M.G. & Burrell, J. (2017). Connected Learning and the Equity Agenda: A Microsociology of Minecraft Play. In Proceedings of the 20th ACM Conference on Computer-Supported Cooperative Work and Social Computing. ACM, New York, NY, 446-457. DOI: <https://dx.doi.org/10.1145/2998181.2998318>
- Anderson, E., Walker, J., Kafai, Y. B., & Lui, D. (2017). The gender and race of pixels: an exploration of intersectional identity representation and construction within minecraft and its community. In Proceedings of the 12th International Conference on the Foundations of Digital Games (p. 17). ACM.
- Basawapatna, A. R., Koh, K. H., & Repenning, A. (2010). Using scalable game design to teach computer science from middle school to graduate school. In *Proceedings of the fifteenth annual conference on Innovation and technology in computer science education* (pp. 224-228). ACM.
- Bluemink, J., Hamalainen, R., Manninen, R., & Jarvela, S. (2010). Group-level analysis on multiplayer game collaboration: How do the individuals shape the group interaction? *Interactive Learning Environments*, 18(4), 365–383. doi:10.1080/10494820802602444
- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of educational research*, 65(3), 245-281.
- Callaghan, N. (2016). Investigating the role of Minecraft in educational learning environments. *Educational Media International* 53, 4 (2016), 244- 260. DOI: <http://dx.doi.org/10.1080/09523987.2016.1254877>
- Cambria, E., & White, B. (2014). Jumping NLP curves: A review of natural language processing research. *IEEE Computational intelligence magazine*, 9(2), 48-57.
- Cipollone, M., Schifter, C. C., & Moffat, R. A. (2014). Minecraft as a creative tool: A case study. *International Journal of Game-Based Learning (IJGBL)*, 4(2), 1-14.
- Clark, D., Nelson, B., Sengupta, P., & D'Angelo, C. (2009). Rethinking science learning through digital games and simulations: Genres, examples, and evidence. Paper presented at the Learning science: Computer games, simulations, and education workshop sponsored by the NAS, Washington, DC.
- Clark, D., Tanner-Smith, E. E., & Killingsworth, S. S. (2016). Digital games, design, and learning: A systematic review and meta-analysis. *Review of educational research*, 86(1), 79-122.
- Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education*, 59(2), 661-686.
- Crook, C. (1994). *Computers and the collaborative experience of learning*. London: Routledge & Kegan Paul.
- Davis, K., Boss, J. A., & Meas, P. (2018). Playing in the Virtual Sandbox: Students' Collaborative Practices in Minecraft. *International Journal of Game-Based Learning (IJGBL)*, 8(3), 56-76.
- De Freitas, S. (2006). Learning in immersive worlds: A review of game-based learning.
- Dezuanni, M., O'Mara, J., Beavis, C., Potter, J., & Gilje, Ø. (2015). 'Redstone is like electricity': Children's performative representations in and around Minecraft. *E-Learning and Digital Media*, 12(2), 147–163. doi:10.1177/2042753014568176
- Dieterle, E., & Clarke, J. (2006). Multi-user virtual environments for teaching and learning. In M. Pagani (Ed.), *Encyclopedia of multimedia technology and networking* (2nd ed.). Hershey, PA: Idea Group, Inc.
- Evans, A. C., Wobbrock, J. O., & Davis, K. (2016). Modeling collaboration patterns on an interactive tabletop in a classroom setting. In ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW '16) (pp. 860-871). New York: ACM Press. doi:10.1145/2818048.2819972
- Fields, D. A., Pantic, K., & Kafai, Y. B. (2015). "I have a tutorial for this": The language of online peer support in the Scratch programming community. In Proceedings of the 14th International Conference on Interaction Design and Children (pp. 229-238). ACM.

- French, D. J., Stone, B., Nysetvold, T. T., Hepworth, A., & Red, W. E. (2016). Collaborative Design Principles From Minecraft With Applications to Multi-User Computer-Aided Design. *Journal of Computing and Information Science in Engineering*, 16(2).
- Gee, J. P. (2008). Learning and games. In K. Salen (Ed.), *The ecology of games: Connecting youth, games, and learning* (pp. 21–40). Cambridge, MA: MIT Press
- Hickey, D.T., Ingram-Goble, A.A., & Jameson, E.M. (2009). Designing assessments and assessing designs in virtual educational environments. *J Sci Educ Technol* 18(2):187–20
- Lave, J. & Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge.
- Lenhart, A., Kahne, J., Middaugh, E., Macgill, E. R., Evans, C., & Vitak, J. (2008). *Teens, video games, and civics*. Washington, DC: Pew Internet & American Life Project.
- Li, M. C., & Tsai, C. C. (2013). Game-based learning in science education: A review of relevant research. *Journal of Science Education and Technology*, 22(6), 877-898.
- Makuch, E. 2014. Minecraft passes 100 million registered users, 14.3 million sales on PC. Retrieved from [www.gamespot.com](http://www.gamespot.com).
- Marklund, B. B., Backlund, P., & Johannesson, M. (2013). Children's collaboration in emergent game environments. In FDG (pp. 306-313).
- Nebel, S., Schneider, S., & Rey, G. D. (2016). Mining learning and crafting scientific experiments: a literature review on the use of minecraft in education and research. *J.Ed.Tech & Soc.*, 19(2), 355.
- Niemeyer, D.J., & Gerber, H. R. (2015). Maker Culture and “Minecraft”: Implications for the Future of Learning. *Educational Media International*, 52(3), 216–226. doi:10.1080/09523987.2015.1075103
- Paraskeva, F., Mysirlaki, S., & Papagianni, A. (2010). Multiplayer online games as educational tools: Facing new challenges in learning. *Computers & Education*, 54(2), 498-505.
- Plass, J., Homer, B., & Kinzer, C. (2015). Foundations of game-based learning. *Educational Psychologist*, 50(4), 258–283. doi:10.1080/00461520.2015.1122533
- Ravitch, S. M., & Carl, N. M. (2015). *Qualitative research: Bridging the conceptual, theoretical, and methodological*. Sage Publications.
- Repenning, A., Webb, D. C., Brand, C., Gluck, F., Grover, R., Miller, S., & Song, M. (2014). Beyond minecraft: Facilitating computational thinking through modeling and programming in 3d. *IEEE Computer Graphics and Applications*, 34(3), 68-71.
- Rogat, T., & Linnenbrink-Garcia, L. (2011). Socially shared regulation in collaborative groups: An analysis of the interplay between quality of social regulation and group processes. *Cognition and Instruction*, 29(4), 375–415. doi:10.1080/07370008.2011.607930
- Short, D. (2012). Teaching Scientific Concepts Using a Virtual World: Minecraft. *Teach.Sci.* 58(3), 55–58.
- Shute, VJ. (2008). Focus on formative feedback. *Rev Educ Res* 78:153– 189. doi:10.3102/0034654307313795
- Steinkuehler, C. A. (2004). Learning in massively multiplayer online games. In *Proceedings of the 6th International Conference of the Learning Sciences (ICLS '04)* (pp. 521-528). New York: ACM Press.
- Tanes, Z., & Cemalcilar, Z. (2010). Learning from SimCity: An Empirical study of Turkish adolescents. *Journal of adolescence*, 33(5), 731-739. doi:10.1016/j.adolescence.2009.10.007
- Thompson, C. 2016. *The Minecraft Generation*. Retrieved from [www.nytimes.com](http://www.nytimes.com).
- Wernholm, M., & Vigmo, S. (2015). Capturing children's knowledge-making dialogues in Minecraft. *International journal of research & method in education*, 38(3), 230-246.

# How Teachers Implement Active Learning: Typologies of Orchestrational Flow

Elizabeth S. Charles, Dawson College, echarles@dawsoncollege.qc.ca  
James D. Slotta, University of Toronto (OISE), jim.slotta@utoronto.ca  
Robert Cassidy, Concordia University, rob.cassidy@concordia.ca  
Michael Dugdale, John Abbott College, michael.dugdale@johnabbott.qc.ca  
Kevin Lenton, Vanier College, lentonk@vaniercollege.qc.ca  
Chao Zhang, McGill University, chao.zhang2@mail.mcgill.ca

**Abstract:** The term, Active Learning (AL) is commonly used in describing pedagogical approaches that engages learners actively in classrooms, with an emphasis on problem solving, inquiry and reflection. While a growing body of evidence supports the effectiveness of such an approach, there is great variation in defining the specific strategies and approaches, making it difficult to advance the field. We examined 19 college instructors for over 220 hours using AL methods. We coded the observed classroom activities according to teacher- and student-centered behaviours. On average, these teachers allocated over 50% of class time to group work. A cluster analysis revealed four distinct patterns of student-centred activity: (1) frequent, short duration; (2) longer duration; (3) less frequent, short duration; and, (4) mixed. Two approaches to workflow were identified: 1) tightly orchestrated and 2) front-loaded with less structured periods of work. Results suggest typologies of instructional patterns, growth trajectories and new directions for examining AL.

## Introduction

Most educators now recognize the need to foster “21st century knowledge skills”, such as critical thinking, collaborative problem solving, and evidence-based reasoning (Hargreaves, 2003; Pellegrino & Hilton, 2012). Science educators have responded to this challenge, exploring new modes of learning and instruction, such as peer instruction (PI), where students in large lecture courses respond to multiple choice conceptual problems using “clickers”, with the tallied results introduced as a powerful mediator of whole class discussions (Crouch & Mazur, 2001). While PI methods hold promise for improving large lecture courses, a movement is now underway to reduce large lectures, emphasizing smaller, recitation-sized sections, led by TAs or instructors. In the “flipped classroom” approach, students spend time at home preparing for class by watching video lectures and reading texts, and class time engaging in active forms of problem solving, small group work, tutorial and recitation (Brookfield, 2012).

Referred to broadly as “Active Learning” (Bishop & Verleger, 2013; DeLozier & Rhodes, 2017), this approach has now engaged many STEM educators, resulting in professional societies (e.g., SALTISE.ca) and university-based centers to support the design of active learning courses. Pioneered by Beichner and his colleagues (e.g., Beichner, Saul, Abbott, Morse, Deardorff, Allain & Risley, 2007), the Student-centered Activities for Large Enrolment Undergraduate Physics, or SCALE-UP method, emphasizes small group tables for student engagement with hands-on activities, physical or digital manipulatives, structured and conceptual problems). SCALE-UP was adapted by MIT for its introductory physics curriculum, known as Technology Enhanced Active Learning (TEAL; Dori et al, 2003), which has provided an important referent for many new initiatives in undergraduate science education. Ruiz-Primo, Briggs, Iverson, Talbot and Shepard (2011) summarize active learning (AL) as comprising four dimensions: conceptually oriented tasks, collaboration, technology, and inquiry based projects. Several studies have measured the benefits of AL (e.g., Dori & Belcher, 2005; Linton, Pangle, Wyatt, Powell & Sherwood, 2014). A large meta-analysis of AL in STEM, shows that exams scores improved by 6% and students were 1.5 times less likely to fail compared with traditional lecture approaches (Freeman, Eddy, McDonough, Smith, Okoroafor, Jordt & Wenderoth, 2014). Moreover, AL has been shown to improve critical thinking, motivation and communication skills (e.g., Kim, Sharma, Land & Furlong, 2013), and increase students’ use of metacognitive processes such as decision making and questioning (Lin, Hmelo, Kinzer & Secules, 1999).

Despite this evidence of success, however, AL remains largely ill-specified and difficult to study with any control (Ruiz-Primo et al., 2011; Brownell, Price & Steinman, 2013). For example, while specific group strategies are often invoked (e.g., group work, gallery walks, collaborative projects, or problem solving) very little definition is provided about the learning processes within those groups, the materials or assessments, or the instructor’s role during the activities (Henderson & Dancy, 2007). Simply naming collaborative strategies fails to

provide sufficient detail about the content, structure or sequencing of activities, nor how they must be configured to ensure efficacy. What makes a hands-on lab activity effective? When should it be used within the AL sequence? How will students collaborate, and to what end? How should design projects be structured, and how should they be assessed? Additionally, AL instruction places new demands on teachers' management and responsibilities for learning, referred to by Dillenbourg (2013) as *classroom orchestration*. Practitioners and researchers require more detail about AL designs, in order to develop a deeper understanding of how to create or adapt them for their own learning designs.

The recent interest in building innovative "Active Learning classrooms" (ALCs) offer an opportunity to gain insight into the instructional patterns used by instructors who purport to engage in AL. Instruction that takes place in these classrooms relies heavily on collaborative learning and student projects, providing opportunities to observe the corresponding teaching practices in a "natural habitat". By examining the nature of teaching and learning in such environments, we can gain an understanding of what instructors are doing and how they integrate AL strategies into their courses. Are there meaningful patterns and/or trends? Are they intentionally designed to produce specific learning and/or instructional outcomes?

We conducted a study of AL "in the wild" to inform our understanding of active learning, and to improve our own approaches to AL design. Our observations and survey of 19 instructors, reported in this paper, can serve as an important referent, informing how we make sense of AL in terms of its patterns of activity structure, discourse and interactions. In many ways, examination of these authentic implementations are akin to what Bereiter (2014) calls for when he described principled practical knowledge, "a type of knowledge that has characteristics of both practical know-how and scientific theory" (p. 5). This research thus aims to inform theoretical perspectives on AL, as well as instructional design and principled instructional practices (authors, 2010).

## Methods

### Design and data collected

This research used a case study design (Yin, 2013), involving mixed methods and ethnographic approach (Denzin & Lincoln, 2000), and part of a larger study involving student group practices and artifact production. The data used included classroom observations consisting of field notes, video recordings, teacher interviews and the pedagogical commitment self-reporting survey - i.e., Post-secondary Instructional Survey (PIPS; Walter, Henderson, Beach & Williams, 2016). PIPS consist of 24 items categorized into student-centred and teacher-centred statements, that indicate a score describing a teacher's practice along the two factors. Two indices were generated from PIPS scores: A student-centred index (SCI) was created from the ratio of the instructor's student-centred score and teacher-centred score; and classification of SC (student-centeredness) based on the SCI such that Level 1 included SCI less than 2, Level 2 included SCI from 2 to 3.5, and Level 3 included SCI greater than 3.5.

### Participants and setting

The study used a purposeful sampling to recruit a total of 19 instructors from three colleges, situated in a metropolitan city in eastern Canada. This sample was selected from a total population of approximately 70 teachers across the three colleges, who voluntarily select to teach their one or more of their courses in active learning classrooms. Their data corpus represents 33 course sections, from 13 courses, and eight disciplines - Physics, Chemistry, Biology, Mathematics, Psychology, History, Humanities and English. Class sizes ranged from 15 to 47 students. All 19 participants had taught in these types of classrooms multiple times, ranging from 5 - 12 semesters (Mean = 9 semesters). Additionally, each institution has some level of support and expectation that teachers using these classrooms will use a student-centred approach and engage with an appropriate professional learning community.

### Procedure and analysis

Observations were conducted by members of the research team, carefully following a common protocol that included field notes and video recordings using GoPro cameras to document the teacher's activity. Multiple class sessions were observed for each teacher, based on an established schedule that set conditions - e.g., accommodation of the teacher preferences, minimum number of classes to observe, timing between observations. In all instances, efforts were made to schedule observations to ensure obtaining a representative sampling of the teacher's practices and their instructional implementations (avg.# of observations = 8.3).

Quantitative data analysis was performed using R (v3.4.1), an open source analytic software environment. Analysis of the classroom observations was performed using StudioCode and a protocol similar to



the COPUS approach (Classroom Observation Protocol for Undergraduate STEM; Smith, Jones, Gilbert & Wieman, 2013). However, our coding sought to elaborate on the types of classroom behaviours during the class session, as well as to document the amount of time spent in each kind of activity. *Capturing the distinct kinds of instructional interaction -- including various forms of lecture, small group, and individual learning activities -- was a primary goal, with the aim of providing a clear account of the frequency, duration, sequence and mix of interactions that occur.* Each class session was coded for presence of the following categories: (1) teacher-centered (lecture/demo); (2) student-centered (group/individual/whole class/student presentation); and, (3) other (administrative work). Actual time spent in each activity was then calculated and recorded for each teacher. We also recorded the type of classroom as being either high-tech “active learning space” (e.g., interactive writable surfaces) versus low-tech “active learning space” (e.g., traditional whiteboards and writing surfaces), in order to investigate whether this variable influenced the nature of classroom interactions.

## Results

Entering into this effort, we recognized that instructors would vary appreciably in how they put the ideas of AL into practice, and for purposes of this paper, we include any student-centered instruction occurring within these classrooms as a form of active learning. A total of 157 observations were obtained, with average class times of 85 minutes, and an average of 8.3 observations per teacher.

We found that instructors spent 59% of their time in student-centred activities (ranging from a low of 24% to a high of 76%), and 40.8% in teacher-centred activities (see Figure 1). A box-plot of these observational data (Figure 2) reveals that student-centred approaches were the primary instructional mode for 15 of the 19 teachers observed. This is in clear contrast to other studies such as Lund et al. (2015) whose sample of 73 teachers spent less than 20% of class time in student-centred activity, with lecture as the primary instructional mode. Even in cases when professional development (PD) training was provided (e.g., Stains, Pilarz & Chakraverty, 2015), lectures remained above 60%, and returned to 80% within two years of the PD intervention. Thus, it is encouraging that AL-oriented instructors exhibit such a clear pattern of student-centred learning designs.

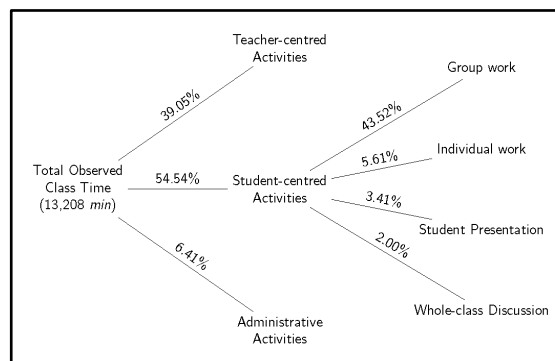


Figure 1. Average proportion of class time spent on different activities by the 19 teachers.

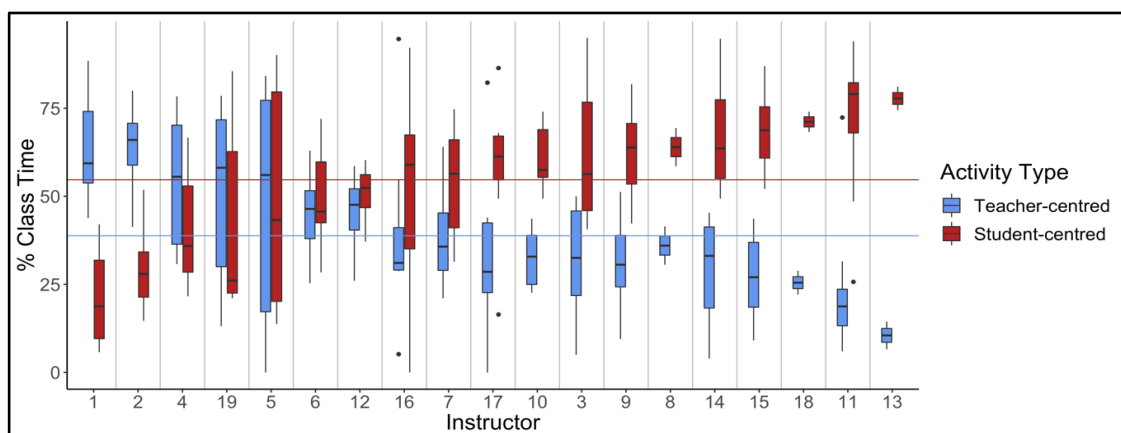
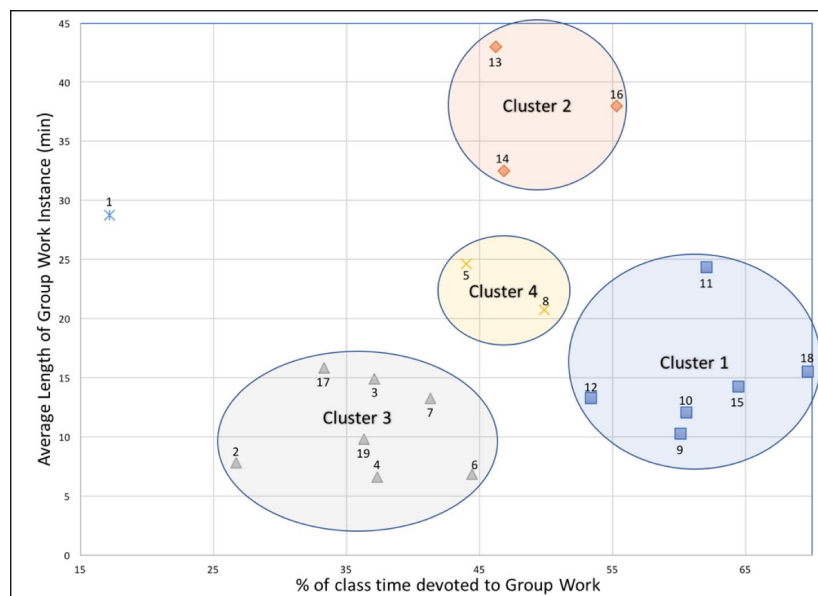


Figure 2. Box plot comparisons of the 19 instructors, ranking by time spent in student centered activities within their observed classes. Horizontal lines are grand means for each activity type.

## What factors impact instructional patterns?

To determine whether the observed patterns of instruction depended on particular characteristics of the teacher or classroom, a MANOVA was performed with activity type (lecture, group work, individual work, student presentation, whole-class discussion) as the dependent variable. Independent variables were teacher, classroom type (high-tech ALC or low-tech ALC), course section, course content and semester (fall or winter session). There were no main effects of classroom type (low-tech or high-tech), course section, or semester, nor significant interaction effects, however there was a significant main effect of teacher, with Pillai's Trace = 1.23 ( $F_{(18, 131)} = 2.37, p < 0.05$ ). This is unsurprising, as different teachers would likely have characteristic instructional patterns, which would give rise to such a significant effect. To further test whether teachers' instructional patterns are stable across contexts, we examined the 7 teachers who were observed multiple times in different contexts. For each of these teachers, a MANOVA was performed to examine whether or not the different contexts led to different instructional patterns. In all but one teacher, no statistically significant difference was found, suggesting that teachers within our cohort use similar instructional patterns regardless of the course context. Post-hoc analysis of the one exception indicated that the difference is attributable to a small dependence of "student presentation" activities in the observed course. This overall lack of variation across contexts suggests that it is safe to aggregate our observations across context for the remaining analyses.

We sought to further understand the characteristics of these active learning designs by focusing on the most commonly occurring form of student-centered activity: group work. One question that could be addressed from our observation data was concerned with the average duration of groupwork. We performed a *partitioning around medoids* cluster analysis (Reynolds, Richards, de la Iglesia & Rayward-Smith, 1992). Findings show that 18 of the 19 teachers fall into one of four clusters (Figure 3). Cluster 1 indicates very frequent but short activity sessions; cluster 2, moderately frequent but long activity sessions; and Cluster 3 shows moderately frequent but shorter activity sessions; finally the two teachers in Cluster 4 employed moderately frequent and moderately long activity sessions.



**Figure 3.** Cluster analysis revealing four clusters of student-centred pedagogy organized by % of class time devoted to group work and average length of group work assigned within class session.

Looking within clusters, we found different orchestrational flows. In Cluster 1 teachers were generally observed as designers of many short duration activities engaging students in collaborative tasks that included group problem solving, peer review and editing, peer instruction (reciprocal teaching), orchestrated by task termination (or interruption) after 10-15 minutes to provide feedback at the class-level. Observations of teachers in Cluster 2 revealed longer group activities with orchestration involving the teacher circling around the classroom, providing feedback to individual groups. Generally, such activities engaged students in project work, sometimes with a distribution of labour and sometimes working jointly to build and/or add to a shared artifact. Cluster 3 teachers orchestrated with shorter activities and fewer of them, appearing to have less experience with

student-centred instruction or teaching in disciplines with lots of content to learn (e.g., biology, history). Much like the earlier cluster, these tasks engaged students in problem solving and peer instruction. Cluster 4 is more difficult to describe because the two teachers appear to engage in different orchestrational flows or styles, teacher 8 being more aware of using group work but hampered because of the student under preparedness (described in interviews).

### What characteristics are shared by teachers within the clusters?

To understand the factors that might account for these clusters, we examined characteristics of the 19 teachers (see Table 1). We found that Cluster 1 teachers were predominantly from a common discipline (physics), as compared with the other clusters, which were more heterogenous. Cluster 1 teachers also generally shared a high commitment to student-centred interactions like small group work, with 4 out of the 6 cluster members at Level 3 on the SCI classification. We examined other factors, such as years of teaching or experience in AL classrooms, but those did not vary systematically (i.e., in any way that could offer explanation for the clusters). Still, in Cluster 3 we recognized that two of the disciplines found there (biology and chemistry) have a strong tradition of focusing on content at the undergraduate levels, particularly as these were introductory courses with multiple sections, where teachers expressed feeling external pressures to prepare students for the common final. This could may explain why their SC index scores are high compared to their actual implementation patterns – i.e., cluster allocation.

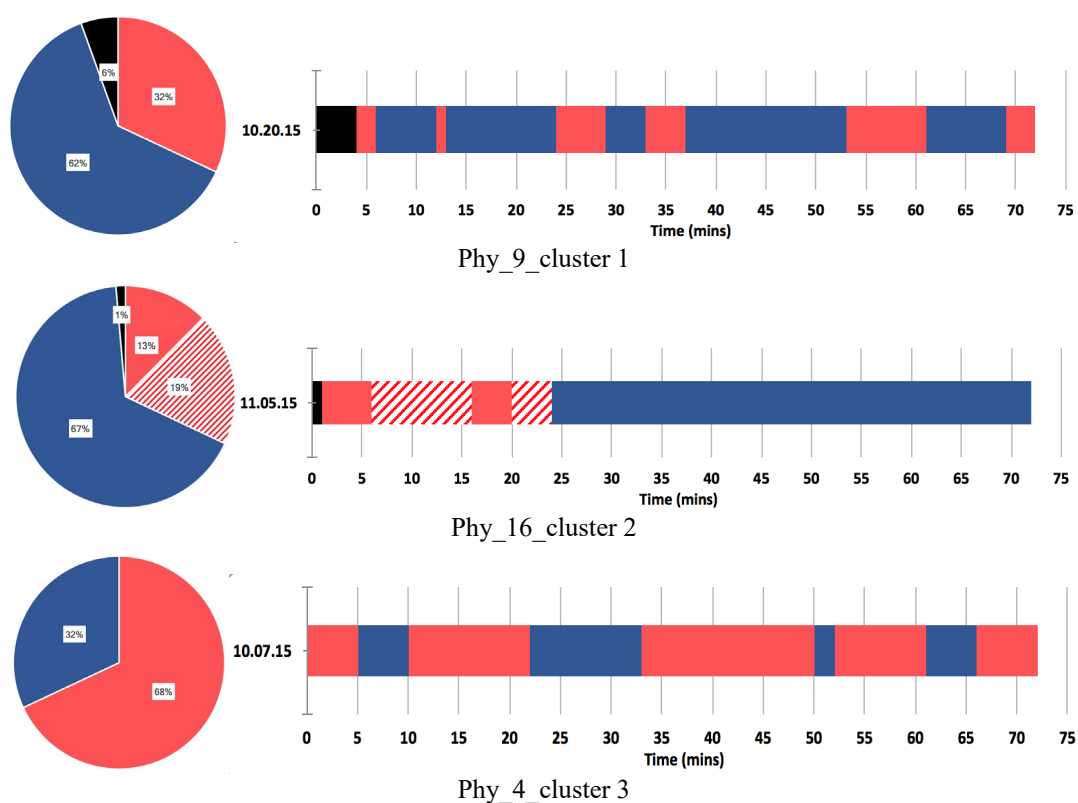
**Table 1: Breakdown of individual teachers information organized by their association to the cluster analysis**

| cluster | Teacher ID* | Teaching experience | AL experience | SCI  | SCI Level | Discipline |
|---------|-------------|---------------------|---------------|------|-----------|------------|
| 1       | 9           | 11                  | 3             | 6.1  | 3         | physics    |
| 1       | 10          | 9                   | 2             | 3.5  | 3         | physics    |
| 1       | 11          | 9.5                 | 2             | 4.5  | 3         | physics    |
| 1       | 12          | 17                  | 3             | 10.4 | 3         | physics    |
| 1       | 15          | 11                  | 2             | 3.4  | 2         | physics    |
| 1       | 18          | n/a                 | n/a           | n/a  | n/a       | math       |
| 2       | 13          | 15                  | 1             | 4.6  | 3         | english    |
| 2       | 14          | 7.5                 | 2             | 2.5  | 2         | sociology  |
| 2       | 16          | 13                  | 3             | 2.8  | 2         | physics    |
| 3       | 2           | 7.8                 | 2             | 1.8  | 1         | chemistry  |
| 3       | 3           | 6                   | 1             | 1.6  | 1         | psychology |
| 3       | 4           | 9.5                 | 2             | 2.0  | 1         | physics    |
| 3       | 6           | 16                  | 2             | 10.6 | 3         | biology    |
| 3       | 7           | 23                  | 2             | 3.1  | 2         | sociology  |
| 3       | 17          | 9                   | 1             | n/a  | n/a       | history    |
| 3       | 19          | 11                  | 2             | n/a  | n/a       | biology    |
| 4       | 5           | 15                  | 2             | 1.8  | 1         | chemistry  |
| 4       | 8           | 10                  | 2             | 5.4  | 3         | physics    |

Notes: Teaching experience, numbers of years teaching; AL experience (proxy used: request to teach in AL classrooms) 1 = low (>2 yrs), 2 = medium (3-5 yrs), 3 = high (> 5 yrs); SCI, student-centred index, see Methods section for details; SCI Level, see Methods section for details. \*Teacher 1 not presented here, b/c not part of cluster analysis.

### Characterizing orchestrational flows

Several studies have focused on teachers' pedagogical implementation, which is sometimes referred to as "pedagogical flow" (Schmidt, 2007), and more recently subsumed within the broader notion of classroom orchestration (Dillenbourg, 2013). Inspired by these notions, we explored the implementation patterns of our cohort of teachers, selecting one teacher from the three main clusters (i.e., Clusters 1, 2 and 3). In an effort to minimize differences in implementation that might be associated with content, we selected the three teachers who teach the same discipline (physics) and specific course (Mechanics), but who fell into different clusters. Because we conducted multiple observations for each of these teachers, and the timeline descriptions for each observation were somewhat distinct (i.e., based on topic of the day and what the students were asked to do), we had to develop some measure of a teacher's "characteristic implementation," which we obtained by selecting the most representative of their respective implementations. This is captured through a pair of graphs – a pie chart to show total proportions of time spent in different forms of interaction (i.e., small group, lecture, or demonstration), combined with a timeline showing the distribution of those interactions over a representative class period. Figure 4 shows the three sample teachers, whose respective pedagogical flow can be seen reflecting the three distinct clusters.



**Figure 4.** The orchestration of AL by three teachers, each representing one of the three main clusters identified. Teacher Phy\_9, is representative of cluster 1, teacher Phy\_16, of cluster 2, and teacher Phy\_4, of cluster 3. Colors, blue=group work; red=lecture; red stripe=demonstration; black=administrative work.

Interestingly, while teachers Phy\_9 and Phy\_16 allocate over 60% of their class time to group work there is a stark difference in their implementations. Phy\_9 has a "stop and go" pattern that is a heterogeneous mix of lecture and student group work – creating a rhythm almost like an up-tempo musical piece. Meanwhile, Phy\_16 shows a two-stage pattern that is "front-loaded" with the teacher lecture and demo followed by a prolonged period of group work. Teacher Phy\_4 illustrates the near inverse of the Cluster 1 teacher. This pattern too is "stop and go" but substantially less time is allocated to group work.

Interviews with these teachers also revealed differences in their instructional objectives and pedagogical commitments. Teacher Phy\_9 stated that he had a pair of aims: (1) to guide students through a series of activities that engage them in exploring the concepts, but also (2) to maintain control over the student's interaction with the content as a means of giving them an opportunity to apply their knowledge but then moving them along. Teacher

Phy\_16 stated that his aim was to allow students to explore the content, and preferred to work with the groups themselves to better understand the individuals and their concerns. The interview with teacher Phy\_4 revealed that while he was also interested in having students try out what they were learning, he was more concerned with moving things along and controlling “the chaos that comes with group work” in active learning classrooms.

## Discussion

Under the assumption that this cohort of 19 college teachers is adequately representative of student-centred teachers, we set out to examine their practice and implementation of AL strategies. Examining the allocations of time to different instructional activity revealed four patterns: high group work with frequent interruptions (Cluster 1), high group work with few interruptions (Cluster 2), moderate group work with interruptions (Cluster 3), mixed (Cluster 4). The Cluster 2 pattern suggested a front-loading of lecture before handing over longer durations of time for students to work on activities. Such implementations appear similar to inquiry-based approaches. On the other hand, Cluster 1 and 3 patterns revealed types of “stop and go” tempos, or what might be considered a heterogeneous mix of lecture and group work – though clearly Cluster 1 is more student-centred.

These identified patterns offer new opportunities to extend the CSCL research on orchestration in empirical ways. For instance, the two types of AL patterns call for very different management of resources and learning – group level versus class level. Additionally, the patterns of Clusters 1 and 3 offer opportunities to explore how and what teachers do in interweaving the lecture and group work and how they integrate AL strategies. The limits of this paper do not allow for the expansion of this investigation but interviews with these teachers reveal clearly different intentions behind these designed implementations. Lastly, the inverse of patterns identified between Cluster 1 and 3 raises the question of whether or not these might not be representative of growth trajectories along the continuum of AL adoption. As such, we might further ask whether or not teachers adoption of this new approach takes on such trajectories and could provide insights on professional development. This study has begun the process of shaping what is to be explored as CSCL and the Learning Sciences begins to consider AL as an instructional practice. We argue that by examining authentic implementations we gain a better understanding of how collaborative learning is orchestrated to meet the needs of integrating group work and lecture, and more importantly, how our findings might be adopted by everyday practitioners. Such investigations also bring us closer to understanding how we might follow up on Bereiter’s challenge of designing with *principled practical knowledge* (Bereiter, 2014).

## References

- Bereiter, C. (2014). Principled practical knowledge: Not a bridge but a ladder. *Journal of the Learning Sciences*, 23(1), 4-17.
- Beichner, R. J., Saul, J. M., Abbott, D. S., Morse, J. J., Deardorff, D., Allain, R. J., & Risley, J. S. (2007). The student-centered activities for large enrollment undergraduate programs (SCALE-UP) project. *Research-based reform of university physics*, 1(1), 2-39.
- Bishop, J. L., & Verleger, M. A. (2013, June). The flipped classroom: A survey of the research. In *ASEE National Conference Proceedings*, Atlanta, GA (Vol. 30, No. 9, pp. 1-18).
- Brookfield, S. D., & Preskill, S. (2012). *Discussion as a way of teaching: Tools and techniques for democratic classrooms*. John Wiley & Sons.
- Brownell, S. E., Price, J. V., & Steinman, L. (2013). Science communication to the general public: why we need to teach undergraduate and graduate students this skill as part of their formal scientific training. *Journal of Undergraduate Neuroscience Education*, 12(1), E6.
- Coburn, C. E., & Penuel, W. R. (2016). Research–practice partnerships in education: Outcomes, dynamics, and open questions. *Educational Researcher*, 45(1), 48-54.
- Crouch, C. H., & Mazur, E. (2001). Peer instruction: Ten years of experience and results. *American journal of physics*, 69(9), 970-977.
- DeLozier, S. J., & Rhodes, M. G. (2017). Flipped classrooms: a review of key ideas and recommendations for practice. *Educational psychology review*, 29(1), 141-151.
- Denzin, N. K., & Lincoln, Y. (Eds.). (2000). *Handbook of qualitative research* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Dillenbourg, P. (2013). Design for classroom orchestration. *Computers & Education*, 69, 485-492.
- Dori, Y. J., & Belcher, J. (2005). How does technology-enabled active learning affect undergraduate students' understanding of electromagnetism concepts?. *The journal of the learning sciences*, 14(2), 243-279.
- Dori, Y. J., Belcher, J., Bessette, M., Danziger, M., McKinney, A., & Hult, E. (2003). Technology for active learning. *Materials Today*, 6(12), 44-49.

- Edelson, D. C., Gordin, D. N., & Pea, R. D. (1999). Addressing the challenges of inquiry-based learning through technology and curriculum design. *Journal of the learning sciences*, 8(3-4), 391-450.
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23), 8410-8415.
- Hargreaves, A. (2003). *Teaching in the knowledge society: Education in the age of insecurity*. Teachers College Press.
- Henderson, C., & Dancy, M. H. (2007). Barriers to the use of research-based instructional strategies: The influence of both individual and situational characteristics. *Physical Review Special Topics-Physics Education Research*, 3(2), 020102.
- Kim, K., Sharma, P., Land, S. M., & Furlong, K. P. (2013). Effects of active learning on enhancing student critical thinking in an undergraduate general science course. *Innovative Higher Education*, 38(3), 223-235.
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational psychologist*, 41(2), 75-86.
- Lin, X., Hmelo, C., Kinzer, C. K., & Secules, T. J. (1999). Designing technology to support reflection. *Educational Technology Research and Development*, 47(3), 43-62.
- Linton, D. L., Pangle, W. M., Wyatt, K. H., Powell, K. N., & Sherwood, R. E. (2014). Identifying key features of effective active learning: the effects of writing and peer discussion. *CBE-Life Sciences Education*, 13(3), 469-477.
- Lund, T. J., Pilarz, M., Velasco, J. B., Chakraverty, D., Rosploch, K., Undersander, M., & Stains, M. (2015). The best of both worlds: building on the COPUS and RTOP observation protocols to easily and reliably measure various levels of reformed instructional practice. *CBE-Life Sciences Education*, 14(2), ar18.
- Pellegrino, J. W., & Hilton, M. L. (2012). *Committee on Defining Deeper Learning and 21st Century Skills*. Center for Education.
- Reynolds, A., Richards, G., de la Iglesia, B. and Rayward-Smith, V. (1992) Clustering rules: A comparison of partitioning and hierarchical clustering algorithms; *Journal of Mathematical Modelling and Algorithms* 5, 475-504 (<http://dx.doi.org/10.1007/s10852-005-9022-1>).
- Ruiz-Primo, M. A., Briggs, D., Iverson, H., Talbot, R., & Shepard, L. A. (2011). Impact of undergraduate science course innovations on learning. *Science*, 331(6022), 1269-1270.
- Schmidt, W. H. (Ed.). (2007). *Characterizing pedagogical flow: An investigation of mathematics and science teaching in six countries*. Springer Science & Business Media.
- Smith, M. K., Jones, F. H., Gilbert, S. L., & Wieman, C. E. (2013). The Classroom Observation Protocol for Undergraduate STEM (COPUS): a new instrument to characterize university STEM classroom practices. *CBE-Life Sciences Education*, 12(4), 618-627.
- Stains, M., Pilarz, M., & Chakraverty, D. (2015). Short and long-term impacts of the Cottrell scholars collaborative new faculty workshop. *Journal of Chemical Education*, 92(9), 1466-1476.
- Walter, E. M., Henderson, C. R., Beach, A. L., & Williams, C. T. (2016). Introducing the Postsecondary Instructional Practices Survey (PIPS): a concise, interdisciplinary, and easy-to-score survey. *CBE-Life Sciences Education*, 15(4), ar53.
- Wieman, C. E., Adams, W. K., & Perkins, K. K. (2008). PhET: Simulations that enhance learning. *Science*, 322(5902), 682-683.
- Yin, R. K. (2017). *Case study research and applications: Design and methods (6th ed.)*. Thousand Oaks, CA: Sage Publications.

## Acknowledgments

This research was supported by a grant from le Ministère de l'éducation et de l'enseignement supérieur, Programme d'aide à la recherche sur l'enseignement et l'apprentissage (PAREA), Quebec, Canada (grant number PA2014-013). And, in collaboration with our colleagues from the SALTISE network <<https://www.saltise.ca/>>.

# Cross-Boundary Interaction for Sustaining Idea Development and Knowledge Building with Idea Thread Mapper

Guangji Yuan, Jianwei Zhang, and Mei-Hwa Chen  
gyuan@albany.edu, jzhang1@albany.edu, mhc@cs.albany.edu  
University at Albany, State University of New York

**Abstract:** This study investigated the cross-classroom interaction in four grade 5 knowledge building communities as the third iteration of design-based research. The four classrooms studied human body systems with the support of the Idea Thread Mapper platform. As the students conducted focused inquiry and discourse within their own community, they posted reflective syntheses of their “Journey of Thinking” to a cross-community space accessible to the broader peers from other classrooms. Students from the four classrooms further participated in a live “Super Talk” to collaboratively address a challenging question. Qualitative analyses of classroom videos, online discourse, and classroom observation notes provide a rich description of how the students interacted within and across communities to develop scientific understandings.

## Introduction

Educational researchers have made substantial advances to develop collaborative learning environments and support collaborative knowledge building among students (Engle & Conant, 2002; Scardamalia & Bereiter, 2006; Slotta, Suthers, & Roschelle, 2014). Further research is needed to extend collaborative interaction to higher social levels and scales (Stahl, 2013), so students can build on the knowledge of other communities across school years for sustained knowledge building (Zhang, Bogouslavsky & Yuan, 2017). This requires designs for cross-boundary interaction and collaboration, which has gained interests in the broader fields (Star & Griesemer, 1989; Star 2013) but still lacks systematic investigation in the field of computer-supported collaborative learning. The purpose of this design-based research is to test and refine designs of cross-classroom interaction for knowledge building with the support of a new collaborative platform: Idea Thread Mapper (Zhang et al., 2015, Zhang et al., 2018).

Existing explorations of cross-community interactions for knowledge building adopted a single layer interaction design. Each classroom gave other classrooms access to their online discourse space so they could read their notes, respond, and supplement with periodic video conferences (Lai & Law 2006; Laferrière et al., 2012). The findings suggest productive classroom changes. Meanwhile, difficulties arose for students to understand other communities’ ideas and discourses without a clear sense of the contexts. New strategies of boundary-crossing support are needed to make knowledge progress accessible across communities.

This research designs cross-community interaction using a multi-level emergent interaction approach, focusing on interactions mediated through boundary objects. “Boundary objects” are artifacts (e.g. reports, tools, models) used to bridge the boundaries (discontinuities) between different social worlds (Star & Griesemer, 1989). Objects from a community often have contextual meanings that are not accessible to other communities. What makes boundary objects effective for bridging different communities of practice is their interpretative flexibility as a “means of translation” (Star & Griesemer, 1989): they have a structure that is common enough to make them recognizable across the different social worlds and allow different communities to interact and work together. Through interacting with shared boundary objects, members from different communities can identify, understand, and reflect on their different practices, leading to enriched understandings within each community and potentially the creation of new, in-between practices (Akkerman & Bakker, 2011).

As noted above, raw distributed online discourse records are hard to use as boundary objects to bridge different knowledge building communities. Therefore, we developed a multi-level emergent interaction approach (Zhang et al., 2017). Students in each community engage in focused inquiry and interactive discourse within their own community’s space. As progress is made, students selectively synthesize fruitful threads of inquiry emerged from their discourse. The reflective reviews and syntheses can facilitate peer learning and build-on within each classroom (Zhang et al., 2018); they may further be shared as boundary objects to enable cross-community interaction. Students from other partner classrooms (or a subsequent student cohort) can use the syntheses of idea threads to view into the progress of inquiry and build connections.

We have conducted a multi-year design-based research to test and refine this multi-level, emergent interaction design (Zhang et al., 2017, Yuan, Zhang & Luo, 2018). The earlier iterations/designs were supported by purposeful uses of existing tools offered by Knowledge Forum (Scardamalia & Bereiter, 2006).

On the basis of the findings, we created and upgraded a new collaboration platform: Idea Thread Mapper (ITM) (Zhang et al., 2018), which interoperates with Knowledge Forum. ITM supports knowledge building interaction in a network of “buddy classrooms.” Students in each classroom co-organize “juicy” wondering areas (shared inquiry foci) as they pursue interactive discourse to deepen their understandings in a domain area. They compose Journey of Thinking (JoT) syntheses to review progress in each area, focusing on (a) overarching topics and problems, (b) we used to think...now we understand... and (c) deeper research is needed. The wondering areas and syntheses are shared in a cross-classroom space that has an easy search tool. By viewing other classrooms’ JoT syntheses, students can get a sense of their inquiry directions and learn from their “big ideas” and deepening questions. As a newly designed feature, ITM allows different classrooms to initiate and participate in live “Super Talk” threads to collaboratively address challenging problems.

This paper reports on a new iteration of this design-based research to refine cross-classroom interaction in a set of four grade 5 classrooms supported by ITM. Our research questions ask: (a) How did students generate JoT syntheses for cross-classroom sharing? (b) How did the four classrooms initiate and pursue “Super Talk” to address challenging problems? And (c) in what ways did the within-classroom discourse and cross-classroom “Super Talk” interact to support deep inquiry and understandings?

## Method

### Classroom design and contexts

This study was conducted in four grade 5 classrooms (with a total of 89 students who were 10-to-11 years old) that studied human body systems over a six-month period using ITM. The four classrooms, labeled as Class 1-4, were taught by two teachers each teaching two classes. Students in each classroom generated interest-driven questions, co-organized wondering areas focusing on various human body systems, and conducted research using various resources. They conducted reflective knowledge-building conversation (called “metacognitive meetings”) in their classroom to build on one another’s questions and ideas while reviewing their progress. The conversation continued on ITM in their online discourse space organized as various idea threads each addressing an overarching problem/theme. As progress is made in each idea thread, students co-created and edited JoT to reflect on their knowledge (Figure 1). The JoT was then shared with all the other classrooms. Drawing upon their knowledge built about the various body systems, students in Class 3 proposed a challenging problem for “Super Talk” across the classrooms: How do people grow? (Figure 3). The proposal was supported by the other classrooms. Students from the four classrooms worked together to discuss this overarching question. Near the end of the unit, each class had a metacognitive meeting to review knowledge gained from the “Super Talk” and build connections with the different human body systems.

### Data sources and analyses

The data sources included classroom observation notes, video records and transcriptions of classroom conversations, and students’ online discourse on ITM and JoT shared in the cross-classroom space. Guided by each of the three research questions, we conducted detailed qualitative analysis of student inquiry process in their home class in connection with student interaction across classrooms. To analyze the quality of students’ JoT syntheses, we conducted content analysis (Chi, 1997) based on the scientific levels (from pre-scientific to scientific) and complexity of ideas (from unelaborated facts to elaborated explanations) (see coding scheme in Zhang et al., 2007). To understand how the Super Talk took place, we read/re-read our detailed field notes and examined the classroom videos to understand the classroom processes. To understand how the within-classroom discourse and cross-classroom Super Talk interacted to support further inquiry, researchers selectively transcribed two metacognitive meeting videos from each class, one from the 5<sup>th</sup> month, and the other from the sixth/last month of the inquiry in which students discussed what they had learned from the Super Talk. We analyzed the classroom discussion to trace the formation of cross-topic connections based on the betweenness centrality measure (Oshima, Oshima & Matsuzawa, 2012).

## Results

### How did students generate syntheses of Journey of Thinking for cross-classroom sharing?

Students engaged in kick-off activities that elicited their interests about the human body in January 2018. Students in each room generated various interests and questions and formulated overarching “wondering areas.” These areas became the focus of the subsequent inquiry by individuals and groups using books, online resources, and models. With inquiry progress made in the next two months, students were introduced to the JoT



function in ITM. Focusing on each area of inquiry, students added reflections on their research individually, and then compiled the reflection entries into whole group JoT (Figure 1). The syntheses of JoT were then shared with all the other classrooms in the cross-classroom sharing space.

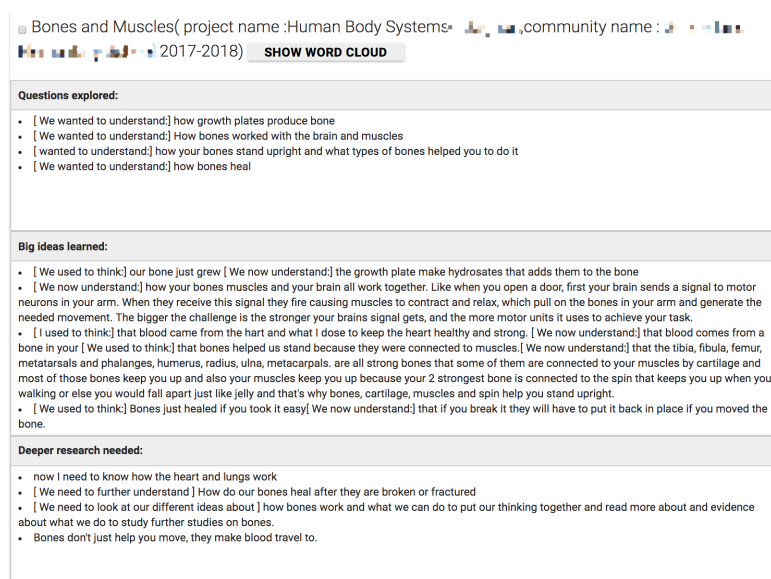


Figure 1. The Journey of Thinking synthesis written for the area of bones and muscles from Class 1.

To examine the quality of students' reflections in their JoT, researchers coded the ideas summarized under "We used to think" and "We now understand" based on scientific sophistication and epistemic complexity (Zhang et al., 2007) (see Table 1).

Table 1: Coding of ideas in the JoT summarized under "We used to think" and "Now we understand" based on data from the current (2018) and the previous iteration (2017) of this research

|                                  | 2018             |                   | 2017             |                   |
|----------------------------------|------------------|-------------------|------------------|-------------------|
|                                  | We used to think | Now we Understand | We used to think | Now we Understand |
| <b>Scientific Sophistication</b> |                  |                   |                  |                   |
| Pre-scientific                   | 33%              | 0%                | 38%              | 0%                |
| Hybrid                           | 50%              | 0%                | 48%              | 0%                |
| Basic                            | 16.7%            | 0%                | 14%              | 18%               |
| Scientific                       | 0%               | 100%              | 0%               | 82%               |
| <b>Epistemic Complexity</b>      |                  |                   |                  |                   |
| Unelaborated Facts               | 83%              | %                 | 95%              | 9%                |
| Elaborated Facts                 | 0%               | 16.7%             | 0%               | 32%               |
| Unelaborated Explanations        | 16.7%            | 0%                | 5%               | 9%                |
| Elaborated Explanations          | 0%               | 83.33%            | 0%               | 48%               |

As Table 1 shows, ideas summarized under "Now we understand" show a higher level of scientific quality and complexity than those summarized under "We used to think." Compared against the data from the previous (2017) iteration of this design-based research in which we tested a similar design without ITM, the ideas summarized under "Now we understand" in the current iteration showed higher levels of scientific and complex quality. A more detailed analysis revealed that the JoT syntheses in 2018 had more words on average than those in 2017 (362.3 versus 170.8 words). These changes suggest the potential improvement made by students in their reflective inquiry with the support of ITM.

## How did the four classrooms initiate Super Talk to address challenging problems?

In the beginning of May, at a metacognitive meeting, students in class 3 noticed the new ITM feature of “Super Talk.” The teacher explained that this function was for all the classrooms to explore big challenging questions and put their knowledge together. Then Class 3 started to discuss possible challenging questions for Super Talk. Three questions were proposed in total: How are all the systems connected? Which two systems are most connected? And why do people grow? A few students reflected on what they knew about how muscles grow, and several other students showed interests in the growth topic as they had grown a lot during the school year. Then they agreed to focus on one topic for the Super Talk, and decided to have to vote for the one that they felt was most challenging and exciting. The topic of “How do people grow?” was selected. This Super Talk topic was proposed and added in ITM and made visible to other classrooms (Figure 2).

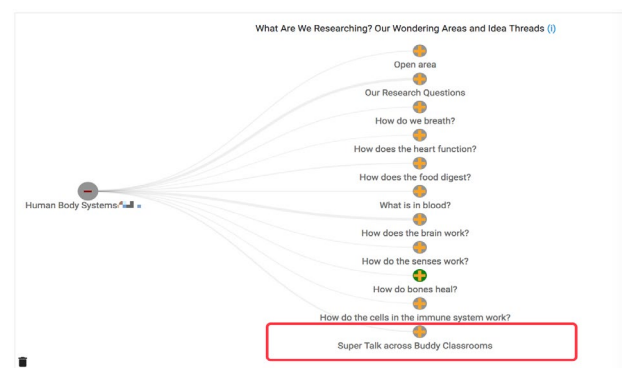


Figure 2. Students’ collective wondering areas about human body systems and the Super Talk area.

In the following week, students from Class 3 first started to contribute knowledge about how the brain, bones and muscles grow drawing upon what they had learned about these body systems. The teachers then advertised the “Super Talk” question to the other classrooms. In each case, the teacher read the notes already posted by Class 3, and discussed what might count as a good note for the Super Talk. In the following week, students from the other classrooms started to build on the ideas in the Super Talk (Figure 3).

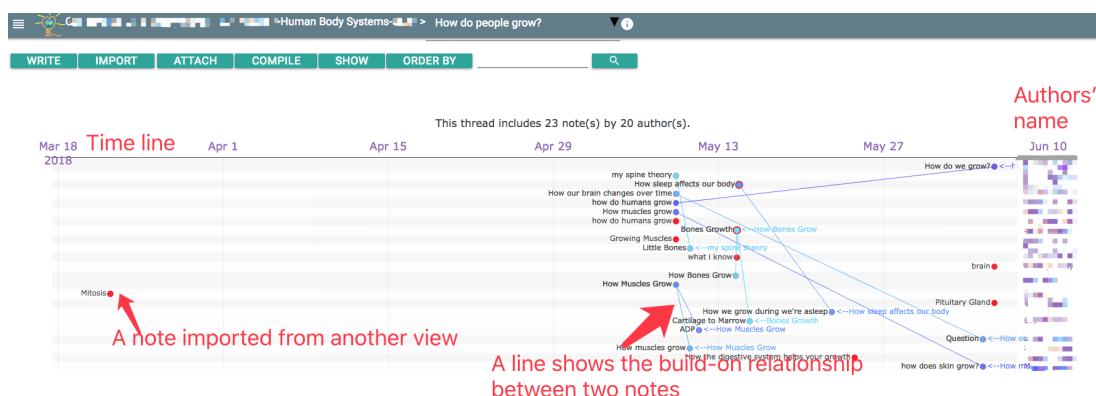


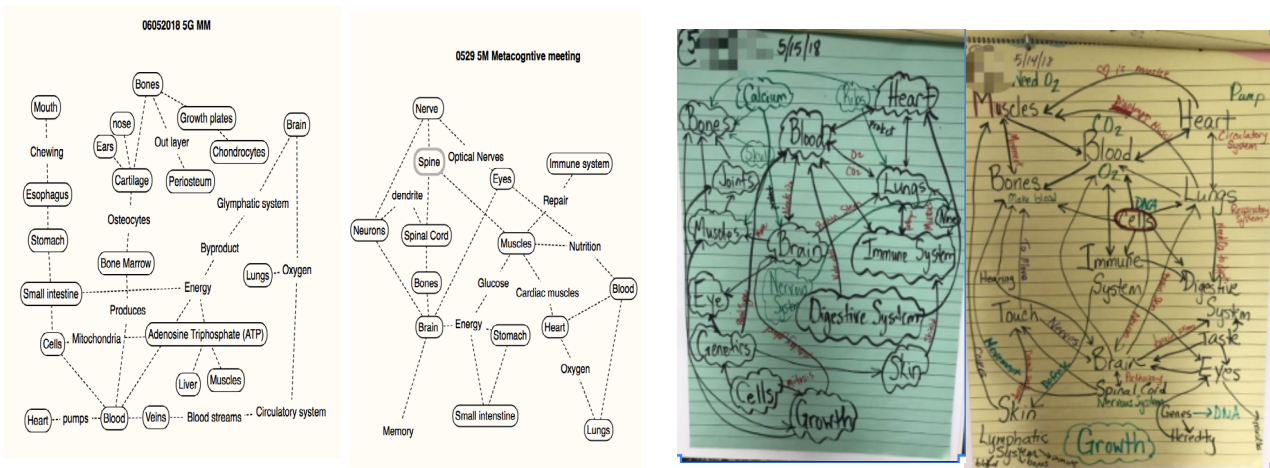
Figure 3. The cross-classroom Super Talk about how people grow. Each dot shows a note, and a line between two dots shows a build-on connection. Each note is positioned based on the date of creation (x-axis) and author (y-axis).

Student A from Class 4, studying the digestive system and energy, found the connection between growth and muscles, so he added more detailed information to the Super Talk. He mentioned that muscles use energy from ATP (Adenosine Triphosphate) for muscle placement. And then Student B from Class 2 built on this idea, saying “Muscles grow by when you stress muscle fibers by lifting heavy weights or doing motions you’re not used to, they rip which lets out a chemical called cytokines which activates your immune system which repairs it bigger than it was earlier which makes your muscles grow. Hypertrophy is how your muscles say you need to work more to make your muscles grow.” At the same time, a new angle of viewing was presented by Student C from Class 1, who wrote about how sleep affects growth as related to her inquiry of the

brain. Her post highlighted that during the Non-Rapid Eye Movement stage (NREM), the body is actually repairing damaged tissues and growing. New detailed information about bones was further expanded by Student D from Class 2 who was studying bones. He mentioned that “bone grows from cartilage, they fuse together and go through a process called ossification.” Later, his peer who studied the same topic built on this note and added more detailed accounts of the process of ossification: “Over time, a different type of cells called osteoclasts head to the middle of the bone to help in. Now, inside osteoclasts there are hydrolytic enzymes and acids. These enzymes and acids will help dissolve the temporarily bone (the cartilage) to make room for the permanent bone (marrow).” Toward the end of the online discussion, Student E from Class 1, who was studying the endocrine system, gave her explanation from the angle of the endocrine system, because the pituitary gland releases a hormone that controls the growth as it plays a huge role in puberty and metabolism. A cross-cutting connection was further built when Student F from Class 2, who was studying cells, found that humans start as cells and all organs are made of cells, and the way cells grow is from mitosis; therefore, cell growth is key to how humans grow. He imported his earlier note about mitosis from his home class discussion into the Super Talk space. This note was read by other students and triggered deeper conversation in other classrooms’ face-to-face meetings.

The Super Talk extended till the end of May with a total of 22 students from the four classrooms participating in the discussion. Students collaboratively explained how people grow involving bone and muscles, brain and nervous systems, cells and genetics, and digestive systems. Approximately 50% of the notes are build-ons, reflecting a higher level of collaborative responses. Student notes were coded based on epistemic complexity (Zhang et al., 2007), with 86% of the notes offering elaborated explanations beyond simple facts.

After students shared their Journal of Thinking and participated in the Super Talk collaboration on ITM, each classroom community continued its inquiry activity in June. The teachers named the last month as the “Month of Connection,” and students extended their learning with the focus on the potential connections between different human body systems. Metacognitive meetings were held to discuss how the different body systems work together, integrating what students had learned from their own inquiry as well as from their peers. As specific body systems and connections were mentioned by students, the teachers took visual notes of the discussion, leading to the co-generation of the concept maps shown in Figure 4.



**Figure 4.** Concept maps of four classrooms metacognitive meetings about connections between human body systems (Class 1, Class 2, Class 3, Class 4, respectively).

During the metacognitive meeting about connections between different human body systems in the 5<sup>th</sup> month, students made solid reflections, build-ons, and had deep conversations about how different systems connect to each other. As their discourse expanded, students started to understand the concept of the cell as the basic unit of the human body, however, they did not mention the mechanism of cells multiplication. The excerpt below shows how students from Class 3, which was taught by Teacher K, tried to understand collaboratively the cell as the start and basic unit of human body systems and the connections between various systems. Below is an excerpt from Class 3’ metacognitive meeting conducted in the 5<sup>th</sup> month of this inquiry.

K21: I think the cells in the immune system should be connected to cells that fight off the germs to fight cells

K8: Actually the immune system is made up cells.

Teacher K: What else is made up of cells?  
 K8: Blood  
 K5: Isn't it everything  
 K8: Yeah, well everything is made up of cells.  
 K5: Well there are all different kinds of cells  
 Teacher K: Ah uh!  
 K1: Yeah, but everything is made up of cells.  
 K8: It's the beginning of everything  
 K5: We could have different specific types of cells, maybe we can go to a bunch of different kinds of cells. Like the brain cells just put it next to the brain?  
 Teacher K: So basically cells is kind of like there is a reason why I drew it in the center. Is it the beginning of every part? Do we have anybody studying cells?  
 K12: I know something about cells. oh, it's that, it's my dad told me, we all started out as one little cell and we start to grow.  
 Teacher K: What? Isn't growth being the big question that we put out up there (on ITM)?  
 K1: Yeah, so we grow taller just by see cells grow in our body  
 K5: and then it just multiply, so there are tons of different types of cells. and that's why there are the differences, because there is a different amount of cells of certain types.  
 K15: I think cells are connected to the growth. because as you grow more cells in your body  
 Teacher K: Every kind of cells  
 K15: If you get a lot of cells in your body at one type, that's how we growth spurts I think  
 Teacher K: I think someone may have some growth expert spurts. Is there anyone want to say the big connections that we are missing?  
 K10: Gene and heredity  
 K8: So the immune system is connected to the digestive system, skin, eyes.  
 Teacher K: Really? Tell us why?  
 K8: So, it is connected to skin cause skin is actually one huge defense against pathogens.

### In what ways did the cross-classroom “Super Talk” leverage deeper discourse in each classroom?

In early June, each class held a face-to-face metacognitive meeting to revisit the question of how people grow. Discourse analysis showed that students brought what they had learned from the “Super Talk” back to their home class discussion and made further connections with the growth of other organs, leading to deepened understandings of the human body systems. Below is an excerpt from Class 3’s metacognitive meeting.

Teacher K: Your brain cells are dying? or not making new ones  
 Student K8: You are not making new ones, but they do start out, they do die as you get older  
 K5: When you run out of brain cells you die?  
 K12: *I saw something on ITM about chromosomes, it is kind of related to growth.*  
 Teacher: What is it, can you reiterate it? What are chromosomes are related to?  
 K7: Mitosis?  
 K5: DNA?  
 Teacher K: Oh, Mitosis? what about DNA? what's that related to?  
 K5: DNA and RNA.  
 Teacher K: What is that related to?  
 K8: RNA is just half of DNA  
 K12: *Mitosis is the process of one cell splitting into two new cells, it is a complex process of many steps. One prophase. In prophase the structures called centrioles move to opposite ends of the cell and fibers come out of them and enclose the cell. And in metaphase chromosomes line up in the center of*

*the cell. Each attaches to two fibers. Chromosome halves pull apart the cell and divide the membrane. Step three is Anaphase and step 4<sup>th</sup> telophase.*

Teacher K: He is talking about really deep science that's behind this (pointing to the drawing) where the one cell is splitting into two equal parts. So when you cut an apple, you know, in the center of the apple gets really cut in half, it really does, that's not the same as what is going on here. With mitosis, it gets cut in half but each half gets exactly the same the central part. Like the same center of the apple grows into both pieces, when it splits apart, and then that's...they split apart to make two identical, and it still has that center of the apple, and what's in the center in the apple, or the center of the cell?

K1: The DNA

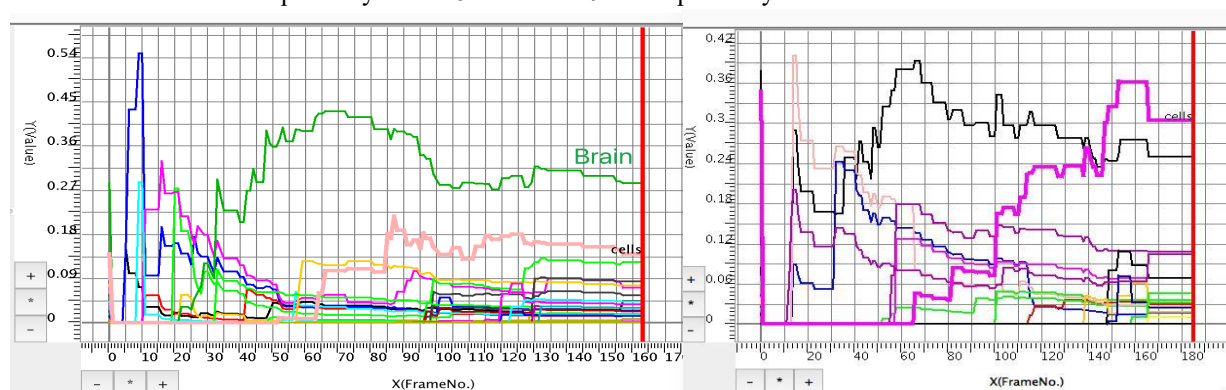
K2: Chromosomes

Teacher: DNA and Chromosomes, and what can you tell us about heredity or DNA

K1: Hair color, eye color.

K17: Your genes there are like the blueprint.

To specifically examine the conceptual connections built between different topical concepts, researchers traced co-occurrence of the main domain concepts (e.g. brain, cell, bones, lungs, and heart etc.) in each conversation turn in the two metacognitive meetings mentioned above before and after the “Super Talk.” Using the network analysis tool KBDEX (Oshima, Oshima & Matsuzawa, 2012), we compared the betweenness centrality of the key concepts, which shows which concepts act as ‘bridges’ between concepts in a network. Figure 5 shows the changing centrality of each key concept over time in each of the two meetings. Before the “Super Talk”, the concept “Brain” stood out as having the highest betweenness centrality among the discussed concepts, suggesting that students’ discourse positioned brain as the central topic connected with other systems. In the meeting after the “Super Talk” activity, the concept of “Cell” had the highest gain in betweenness centrality. Student K12, who acted as broker, brought back the concept of cell mitosis from the “Super talk” and triggered extended discussion related to cells in the home class. According to the science standards, the concepts of cell and mitosis are required by Grade 8 and Grade 9-12 respectively.



**Figure 5.** Betweenness centrality of the key concepts discussed in the metacognitive meeting before (Left) and after the “Super Talk” meeting (Right).

## Discussion and conclusions

This study tested a multi-level emergence design of cross-community interaction, with students engaging in focused inquiry and discourse within their own classroom while generating reflective JoT syntheses as boundary-crossing objects, and participating in the “Super Talk” to investigate a challenge problem together. The results provided an elaborated account of the processes to generate reflective syntheses and pursue Super Talk. Students showed solid reflections in JoT with the help of ITM compare with last year and students built on knowledge building interactions through the cross-classroom collaboration to extend and enrich the discourse in their home class. More specifically, with collaboration going on in the cross-boundary space, students further expanded their social connections, engaged in addressing challenging research questions and demonstrated higher level collaboration and understanding beyond their classrooms’ walls. The cross-community collaboration also enabled students to bring back valuable insights to their home class and triggered further build-on and leveraged the home class’ understanding to “rise above” (Scardamalia& Bereiter, 2006). Cross-classroom interactions offer a multi-layer structure which enable the information flow across various sites in a

mutually understanding approach that suggested by Stahl (2013). The analysis of the classroom discourse reveals the benefits of a higher level of interaction enabled by the updated design of ITM as a knowledge infrastructure. Expanding our previous year's work (Zhang et al, 2017. Yuan et al., 2018), the findings shed light on the possible designs and processes to enable collaborative knowledge building across a network of classrooms in a larger learning environment and ongoing learning process. These research results also provide insights that may guide teachers' adoption of cross-classroom collaboration to support student inquiry and extend knowledge building. During the learning process, the teachers played a crucial role in guiding students' cross-classroom collaborations, embracing the ideas from other communities to extend students' thinking, and using child-friendly language to leverage students' understanding, and making connections with other topics in the current classroom.

## References

- Akkerman, S. F., & Bakker, A. (2011). Boundary crossing and boundary objects. *Review of educational research, 81*(2), 132-169.
- Chen, J., & Zhang, J. (2016). Design Collaborative Formative Assessment for Sustained Knowledge Building Using Idea Thread Mapper. In: Proceedings of the International Conference of the Learning Sciences (ICLS 2016). Singapore: International Society of the Learning Sciences.
- Chi, M. T. H. (1997). Quantifying qualitative analysis of verbal data: A practical guide. *Journal of the Learning Sciences, 6*, 271-315.
- Engle, R. A., & Conant, F. R. (2002). Guiding principles for fostering productive disciplinary engagement: Explaining an emergent argument in a community of learners classroom. *Cognition and instruction, 20*(4), 399-483.
- Laferrière, T., Law, N., & Montané, M. (2012). An International Knowledge Building Network for Sustainable Curriculum and Pedagogical Innovation. *International Education Studies, 5*(3), 148-160.
- Stahl, G. (2013). Learning across levels. *International Journal of Computer-Supported Collaborative Learning, 8*(1), 1-12.
- Scardamalia, M., & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 97-115). Cambridge University Press.
- Scardamalia, M. (2004). CSILE/Knowledge forum®. *Education and technology: An encyclopedia*, 183-192.
- Slotta, J., Suthers, D., & Roschelle, J. (2014). CIRCL primer: Collective inquiry and knowledge building. *CIRCL primer series*.
- Star, S. L., & Griesemer, J. R. (1989). Institutional ecology, "translations" and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science*.
- Oshima, J., Oshima, R., & Matsuzawa, Y. (2012). Knowledge Building Discourse Explorer: a social network analysis application for knowledge building discourse. *Educational technology research and development, 60*(5), 903-921.
- Yuan, G., Zhang, J., & Luo, C. (2018). Cross-Classroom Interaction Among Four Grade 5 Knowledge Building Communities. Conference on American Educational Research Association. New York.
- Zhang, J., Bogouslavsky, M., & Yuan, G. (2017). Cross-Community Interaction for Knowledge Building in Two Grade 5/6 Classrooms. International Society of the Learning Sciences.
- Zhang, J., Chen, M.-H., Tao, D., Lee, J. Sun, Y., & Judson, D. (2015). Fostering Sustained Knowledge Building through Metadiscourse Aided by the Idea Thread Mapper. International Conference on Computer Supported Collaborative Learning (Vol. 1, pp. 166-173).
- Zhang, J., Scardamalia, M., Lamon, M., Messina, R., & Reeve, R. (2007). Socio-cognitive dynamics of knowledge building in the work of nine- and ten-year-olds. *Educational Technology Research and Development (ETR&D), 55*(2), 117-145.
- Zhang, J., Tao, D., Chen, M.-H., Sun, Y., Judson, D., & Naqvi, S. (2018). Co-organizing the collective journey of inquiry with Idea Thread Mapper. *Journal of the Learning Sciences, 27*, 390-430.

## Acknowledgments

This research was sponsored by the Cyberlearning Program of National Science Foundation (#1441479). The views expressed in this article are those of the authors and do not necessarily reflect the views of the funding agency. We thank our collaborating teachers and their students for their creative classroom work enabling this research.



# Embodiment of Computational Thinking During Collaborative Robotics Activity

Theodore J. (TJ) Kopcha, Ceren Ocak  
tjkopcha@uga.edu, ceren.ocak@uga.edu  
University of Georgia

**Abstract:** This case study explored how young children's thinking was embodied during collaborative robotics activity. Two 5th grade learners were recorded as they worked together to program and navigate a robot across a 3'x3' grid of obstacles. The video was then analyzed using a grounded theory approach (Glaser & Strauss, 1967) to determine how cognition manifested in an embodied manner (e.g., gesturing, use of conceptual metaphor) through the tools and affordances employed by the learners. Four primary tools were used by learners in solving their robotics task: (1) basic robot moves, (2) gestures, (3) the map and the computer, and (4), the grid and the robot. These tools provided symbolic access to direct experiences, supporting an emergent and reactive process in which perception and action were intimately linked. Implications for computer science and improving the design of computer-supported collaborative learning are discussed.

## Introduction

Educators and policymakers currently are interested in integrating computer science into K-12 curriculum. At the heart of computer science is computational thinking, a problem-solving process in which both abstract and automated processes are involved in formulating problems and developing solutions during programming activity (Cuny et al, 2010; Wing, 2006). Learners who engage in computational thinking decompose the larger task into smaller subgoals, use algorithmic thinking and pattern recognition to create new and novel programs that address those subgoals, and debug programs that are not performing optimally (Yadav, Mayfield, Zhou, Hambrusch, & Korb, 2014). These skills are important for developing a competent 21st century workforce, making computational thinking in K-12 settings a critical and current issue (Israel, Pearson, Tapia, Wherfel, & Reese, 2015; Sengupta, Kinnebrew, Basu, Biswas, & Clark, 2013).

One way computational thinking and computer science are brought into K-12 classrooms is through robotics activities. Robotics creates an opportunity for learners to work collaboratively to program a robot and complete various tasks (Yuen et al., 2015). The physical nature of this activity is important. Robotics activity creates an opportunity for learners to make use of various tools and their affordances (e.g., the robot) to connect their abstract thinking with the environment. Thus, learners who engage in robotics activity are likely to experience an embodied experience, where mind, body, and environment are intimately connected (Gibson, 1979) in an effort to set and achieve specific goals with the robot.

However, little is known about how young children's thinking might be embodied during robotics activity. Research on computer science and programming is lacking from K-12 classrooms (Lye & Koh, 2014). Few if any studies of computational thinking examine the process through the lens of embodied cognition (Grover & Pea, 2013). Likewise, there is a need for empirical studies of young children who engage in robotics activity; early success with computer science through robotics has the potential to increase interest in and success with STEM outcomes in future coursework (Sullivan & Bers, 2016).

The purpose of this case study was to explore how young children's thinking was embodied during collaborative robotics activity. Two 5th grade learners were recorded as they worked collaboratively to program and navigate a robot across a 3'x3' grid of obstacles. The video was then analyzed using a grounded theory approach (Glaser & Strauss, 1967) to determine how cognition manifested in an embodied manner (e.g., gesturing, use of conceptual metaphor) through the tools and affordances employed by the learners. This study provides key insight into the ways that learners learn together through the help of digital technologies; examining computational thinking as an embodied phenomenon holds implications for collaborative robotics activities with young learners.

## Overview of ecological embodiment in its implications for collaboration

### Ecological perspective towards embodiment

Gibson's ecological psychology (1979) laid the foundation for our research. An ecological perspective towards embodiment suggests that the body serves as a mediator between perception and action; both perception and action are influenced by the environment (Gibson, 1977, 1979). From this perspective, thought is "a whole-body activity in context than simply an in-the-head process" (p.171). In other words, embodied cognition involves a complex relationship between the mind, body, and the environment. This stands in contrast with our understanding of information processing, where perception and action are treated "separately and sequentially" (Michaels & Palatinus, 2014, p.20). As argued by Richardson, Shockley, Fajen, Riley, and Turvey (2009), perception and action are mutual and act together to meet a particular goal. This suggests that thinking is a goal-directed reaction against what is perceived from the environment (Chemero, 2011, Michaels & Palatinus, 2014). In this paper, we draw on the basic premise of ecological psychology to explore children's thinking as a whole-body activity, where the environment guides their decision making.

## Affordances

The concept of affordances is at the forefront of the ecological perspective. Affordances are conceptualized as opportunities or possibilities for action placed in the environment (Rietveld & Kiverstein, 2014). For example, "a flat horizontal surface affords standing and walking, a graspable rigid object affords throw" (Hirose, 2002, p.290). That being the case, the notion of affordance has been gaining prominence in different disciplines such as game design and development, computer science, and learning sciences (Harlow, Dwyer, Hansen, Iveland, & Franklin, 2018; Young, 2013). For instance, Linderoth (2012) used the perception of the affordances in the games and acting on them to understand the gameplay. Hammond (2010) examined the concept of affordance regarding its implications for the use of information and communications technology (ICT) in teaching and learning.

Tools can also serve as an affordance from the perspective of ecological embodiment. Liben (2002) explained how "tools are the functional extension of the environment, it has specific affordances and provides new opportunities for action" (p. 290). Tools can amplify perception to move beyond the physical capabilities of the body. In addition, tools provide opportunities for learners for "active material engagement" (Malafouris, 2013, p. 169). Therefore, the tools become an affordance as the user acts upon them. In this way, the tool and the user of the tool are often considered as a single unit of analysis (see Harlow et al., 2018). That is, the user and the tools become a compound element of a larger system in a cycle of perception and action.

## Embodied collaboration and gestures

Physical gestures are considered strong evidence of embodied cognition (Alibali et al., 2014). Gestures also reveal the features of collaboration in that "gestural communication is based on a common conceptual ground, on the shared knowledge of what 'we both know together'" (Vasc & Ionescu, 2014, p. 150). Vasc and Ionescu (2014) further established that collaboration in a given medium has two sides: the receiver and the communicator; the receiver knows the meaning that a gesture carries and the communicator uses the gestures with confidence that the receiver would understand since it emerges out of a shared experience. That being the case, learners' gestures add another layer to understand how collaboration emerges as an embodied phenomenon as pairs of learners mutually act on the affordances they perceive. In this study, we therefore examine the way tools were used as affordances, and how gesturing supported learner's thinking as they engaged in collaborative problem solving during robotics activity.

## Methods

The video data that was analyzed in this study was collected as part of a larger design research project that focused on developing and testing a robotics unit. The unit was designed as integrative STEM curriculum that supported computational thinking and embodied cognition during collaborative problem solving (see Kopcha et al., 2017). Six 5th grade teachers from a local school setting played an integral role in developing and testing the unit. The unit was enacted over a 2-week period, where learners worked in small groups to navigate a robot through a 3'x3' grid containing various obstacles. As part of the unit, learners explored science and math-related concepts that met specific state standards, including: constructive and destructive forces; operations with fractions; coordinate grid; and algebraic thinking.

Video of Learner activity was recorded in each of the six teacher's classrooms throughout the three class periods in which learners engaged in programming the robot (120 minutes total). Cameras were placed such that small group activity was recorded at both the computer (i.e., when programming) the robot as well as the grid where they executed their program. This provided data for fine-grained analysis of gesturing as well as evidence of the different forms of computational thinking (e.g., problem decomposition, debugging).



To investigate *embodied cognition during collaborative robotics activity*, we focused on the final 60-minute segment of video in which learners completed the majority of the robotics task. We reviewed the video data from each classroom and ultimately selected a single case to explore our research question. The case was a pair of 5th grade learners (one boy, one girl) who successfully completed the robotics activity. Using a single case allowed for an in-depth and fine-grained analysis of the how the learners used each tool and affordance to successfully complete the robotics task. It also allowed us to better understand the ways that various tools served as an affordance to support successful robotics activity as an embodied collaborative experience.

## Instructional context

The robotics task was designed around the source-path-goal schema (Johnson, 2008). Learners were tasked with helping scientists use a robot to collect three samples from an active volcano. These three samples served as the checkpoints that the robot needed to visit in the overall task. Independent from the what path it followed, Learners were asked to program the robot to stop at each collection then arrive back to its original location. The pair then collaborated to determine an overarching solution path and the decompose the overarching task into smaller subunits.

The active volcano was represented on a 3-by-3-foot grid consisting of 36 six-inch squares. The three collection sites were scattered about the grid, along with a variety of elements associated with an active volcano (e.g., dust plume, flowing lava). Learners were challenged to maneuver the robot around the grid, avoiding the elements of the active volcano and visiting each of the three collection sites. The grid served as a testing site where learners practiced the codes that they wrote.

Computer-supported collaboration took place throughout the activity. Learners were put into pairs and tasked with breaking the task down into sub-goals, programming the robot, and accomplishing as much of the overall task as possible. They repeatedly moved between programming the robot at the computer and testing their programming at the grid (see Figure 2). With every attempt, learners continuously advanced and optimized their calculations to be able to move the robot in the coordinates of the grid precisely.



Figure 1. The visual block coding program (left) and the grid representing an active volcano (right).

## Data analysis: A grounded approach

Knowing that gestures carry a semantic value (Alibali, 2005), we transcribed the video and audio data simultaneously for the selected case. Transcripts included the full conversation between learners, paired with short sequences of images from the video to illustrate specific gesturing and learner action. We then followed a grounded theory approach (Glaser & Strauss, 1967), employing an inductive data analysis process to generate our own theoretical understanding of the ways thinking was embodied during robotics activity. A grounded theory approach served the aims of this study because it results in identifying overall patterns that are grounded in data.

Data analysis started with open coding (Johnson & Parry, 2016). During open coding, we tried to be “stay close to the data” (Charmaz, 2006, p.49) by avoiding apriori theoretical presumptions that we had. Our initial codes included: tools/affordances, gestures, and semantics/verbal content. Tools/affordances included the objects in the environment that the learners drew upon as part of their collaborative problem solving, such as the map and the computer and the robot at the grid. Gestures were any hand or body movements made by the learners during collaborative work; this often took the form of imitating the movement of the robot. Semantics/content included specific aspects of note such as descriptions of specific gestures (e.g., learner uses hands to mimic wheels of robot turning right) or the meaning behind those gestures. These codes were broad enough to begin analysis while leaving room for new codes to emerge.

After the open coding, the researchers came together to identify a core variable from the initial codes that was also related to our inquiry. Our goal was to better understand the tools that were used as affordances by

learners and how their thinking was embodied with and through those tools. Our core variable was therefore the tools that learners drew upon during collaborative problem solving. After identifying our core variable, we returned to the transcripts to engage in selective coding and establish the relationship between the categories surrounding the core-variable (Johnson & Parry, 2016). Using the participants' own words to examine these relationships helped retain the essence of the mutual conversation and understand the affordances and semantic context that framed the gestures (Birks & Mills, 2011). This confirmed that the learners used four primary tools while solving their robotics task (Charmaz, 2006): the robot, the grid, the map and the computer.

Two additional coding categories emerged from this deeper analysis. First, we coded basic robot moves that the learners developed as part of a shared language during their problem solving, such as turning at a right angle and moving one unit square on the grid. Second, we coded specific gestures that also served as a fundamental tool for communicating problem solving. These two codes were strongly related in that the gestures used for problem solving often reflected basic robot moves. This served as an indicator that learners' thinking was distributed onto "external information bearing structures" (p.15) (e.g., hand motions indicating turns or unit movements) to articulate their underlying cognition and improve their problem-solving during computational thinking (Rowlands, 2010).

## Findings

Four primary tools that the learners used while solving their robotics task were: (1) basic robot moves, (2) gestures, (3) the map and the computer, and (4), the grid and the robot. These tools illustrate the ways that thinking was embodied during their collaboration.

### Tool 1: Basic robot moves

The learners began by establishing an overarching solution path. In this case, the overarching solution was to program the robot to visit the three collection sites without hitting the obstacles scattered in the grid (see Figure 2). The learners then formulated that path into two basic robot moves: (1) a unit square and (2) a 90-degree turn. The unit square was a move that traversed one of the squares on the grid. It was created explicitly by the pair to solve the problem; units could be combined to represent larger movements by the robot. In this way, these basic robot moves represented an abstraction of the 6x6 grid in that they represented a group of underlying algebraic thinking - moving one unit meant making the robot move at a specific speed for a specific amount of time. If two units were desired, the learners would double the time of associated with that speed. This increased the precision of their calculations while allowing them to work more effectively without repeated visits to the grid.

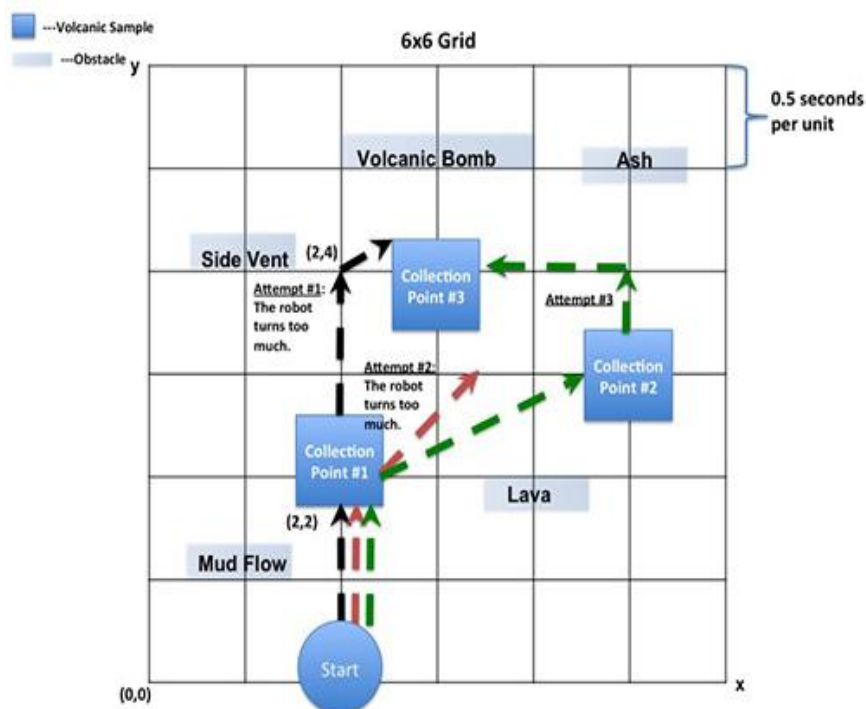


Figure 2. The dotted lines represent the learners' different attempts at completing the robotics tasks.

The 90-degree turn represented an abstraction of multiple underlying algebraic calculations, in the same way as the unit square functioned. To achieve the 90-degree turn, the pair needed to think about multiple variables at once – the relational speed of the wheels with one another as well as their turning direction and duration of the movement. Negotiating those variables led to the development of a shared language in which each of the basic moves (i.e., unit square and 90-degree turn) made collaborative problem solving more efficient and effective. For example, the learners repeatedly returned to the computer after testing a potential set of movements from the robot. Here, they re-enacted the robot’s movements at the grid and constructed new movements to correct any problems with their solution path. They accomplished this by invoking these basic moves to articulate their thinking (e.g., “We need to go forward one, then turn right.”)

## Tool 2: Gestures

The findings from the data indicated that the emergent physical gestures supported learner’s thinking as they engaged in collaborative problem solving during robotics activity. The gestures largely served as a re-enactment of the robot’s movement at the grid. The gestures helped the learners gain the perspective of the robot, specifically when debugging the solution and reprogram the robot’s movement. In this way, their gestures served as a tool for thinking through and communicating their ideas for combining basic moves into more complex movements (see Figure 3).

The most common combination of basic moves was moving straight and turning; this took place five times over the course of the entire solution path. As the learners collaboratively planned these combinations of basic moves, they employed several gestures to denote the movement of the robot. Figure 3, parts a, b, c, and d, display the gestures associated with this sequence of movements: “We stopped right there and we turned to the left [3a]. And now we need to go forward for one second at speed 5 [3b].” The boy then moves his hands apart to match the distance he wants the robot to traverse after turning (3c). The girl then programs the sequence in the computer as the boy begins to repeat his thinking (3d). The type of gestures reveals how learners built upon their basic moves in an embodied fashion to work more adeptly between their abstract and concrete thinking. Lakoff and Nunez (2000) explained how, as humans, we naturally draw on our basic bodily movements to better understand abstract concepts. From this perspective, the learners’ gestures suggest that their thinking was highly embodied as they planned a solution path for the robot.



Figure 3. Gestures communicating basic moves combined into larger sequences of movement and thinking.

## Tool 3: The map and the computer

The learners combined the visual block coding (i.e., computer program) and a small, hand-drawn map to serve as a tool for supporting their collaborative problem-solving. The map and computer together served as a physical representation of their thinking; they would visualize movements from the code itself and then attempt to act those movements out to determine if they would be effective. This is strong evidence of embodiment - learners were offloading cognition onto the map and computer, using the environment to mutually inform their cognition (Wilson, 2002). This is important because it shows how learners dealt with the abstract aspects of computer programming.

Figure 4 displays how the boy used the map (Figure 4a) and computer (4b) to plan a 90-degree turn. The boy explained: “That is where we take the turn (4c). And, that is delay [pointing to the screen, he finds the code that represents the delay] (4d).” The boy’s gestures suggest that he has found two physical representations of his abstract thinking to better communicate with his partner. The learners are looking at both the codes on the screen and the location on the map to predict how the robot will move. They continued:

B: To take the right turn, we need to set the time.

G: For how many seconds?

B: The half, point five,

G: Point five... Sure [pointing to the screen, both learners look at the monitor and come to a mutual decision].

B: And now we need to set the delay.

This exchange suggests that the tools served as an affordance that helped them come to agreement. The physical representations of both the grid and the movement of the robot helped the learners visualize their thinking and determine whether they agreed with each other.

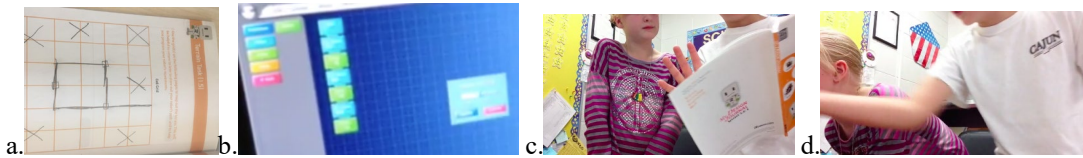


Figure 4. The map and computer together served as a physical representation of abstract thinking.

#### Tool 4: The robot at the grid

The robot also served as an important affordance that mediated perception and action. As the learners viewed the robot at the grid, they watched to see whether the robot's movements aligned with their overarching solution path for completing the task. To the extent that those movements did not, the learners returned to their other affordances (e.g., gestures; the map and computer) to debug their program or revise their intended goal in a collaborative fashion. This repeated itself until the task was successfully completed.

Figure 5 displays this process. At the grid, the learners see how the robot did not reach the intended destination (Figure 5a). The girl identified the problem: "We need to go really faster. Not alone faster but ... [pause, waiting for a response from her partner] (5b)." The boy then responds: "Let's make it the time [that I've used] before [waiting for an approval from the partner] (5c)." The girl then offers a guess: "Yeah, make the time...so point 3 more?" The boy then recognizes how they tried this speed setting previously, leading to a new solution in which they retrace their previous attempts at completing the overarching solution path.



Figure 5. The grid and the robot serve as a mediator between perception and action.

### Conclusion and implications

This study presents a grounded theory of problem solving during robotics activity from an embodied perspective. The learners engaged in a process in which they were repeatedly evaluating whether their current thinking helped them achieve their overarching solution path. This process was emergent and reactive - sometimes they debugged their programming while other times they revisited their solution path. This dynamic process emphasizes the nature of cognition from the perspective of ecological psychology -- the learners set a goal, and information from the environment confirmed whether their actions helped achieve that goal. This is particularly apparent when looking across the learners' attempts at solving the overall task. Their second attempt took them in an entirely different direction than their first (see Figure 1). They then reacted to what emerged within the environment. Their third attempt is a continuation of their second rather than a re-creation of their first. In this way, their thinking shaped the environment while, at the same time, the environment shaped their thinking. This dynamic interplay between mind, body, and environment is at the heart of embodied cognition and typifies Gibson's (1979) notions of an affordance.

The tools identified in this study, then, illustrate how collaborative robotics activity is an embodied experience. Liben (2008) explained how, as humans, we engage in abstract thinking by manipulating external spatial representations both directly and symbolically. In this study, the robot at the grid provided direct access to the learners' problem solving. The learners they drew on their other tools to create symbolic access in the absence of direct access - they used gestures, the map, and the computer to communicate their thinking and create an effective solution.

These findings hold several implications for the embodiment of computational thinking during collaborative robotics activity in K-12 settings. To begin, learners may need a variety of tools to support their engagement in abstract thinking. Tools that provide symbolic access to their direct experiences with the robot can help them offload cognition into the environment during complex problem-solving. This can support collaborative activity. In this study, creating symbolic access helped the learners make their abstract thinking visible to each other. Tools such as maps, block programming, and basic movements helped them more effectively decompose a task, set goals, and create solutions that are mutually agreed upon. Others have similarly found that these tools can support robotics activity in the classroom (Harlow et al., 2018; Liben, 2012).

Additionally, this study offers insight into the way that collaborative robotics activity is an embodied phenomenon. While a great deal of attention has been spent on robotics and computer science, few if any studies view it from an embodied perspective. Viewing this activity as an embodied phenomenon opens a wealth of unexplored questions and potential for future research. For example, how can we design learning environments from an embodied perspective and be more responsive to learners' needs? How does the design of a robot affect how learners move between symbolic and direct access, and can we reduce the distance between these types of access? Moreover, what is gained by studying how we solve complex problems not as a linear, sequential process but rather as a coupling of perception and action in an emergent, complex setting? This study is an initial attempt at exploring these questions in the context of computer science and improving the design of computer-supported collaborative learning in today's schools.

## References

- Alibali, M. W. (2005). Gesture in spatial cognition: Expressing, communicating, and thinking about spatial information. *Spatial Cognition and Computation*, 5(4), 307-331.
- Alibali, M. W., Nathan, M. J., Wolfgram, M. S., Church, R. B., Jacobs, S. A., Johnson Martinez, C., & Knuth, E. J. (2014). How teachers link ideas in mathematics instruction using speech and gesture: A corpus analysis. *Cognition and instruction*, 32(1), 65-100.
- Birks, M., & Mills, J. (2011). Essentials of grounded theory. *Grounded theory: a practical guide*, 11-26.
- Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*: Sage.
- Chemero, A. (2011). *Radical embodied cognitive science*: MIT press
- Cuny, J., Snyder, L. & Wing, J.M.: Demystifying computational thinking for noncomputer scientists. Work in progress (2010)
- Gibson, James J. (1977). "The Theory of Affordances." In Robert Shaw and John Bransford (eds.), *Perceiving, Acting, and Knowing*. Hillsdale, NJ: Lawrence Erlbaum Associates
- Gibson, J. J. (1979). The ecological approach to visual perception. Boston: Houghton Mifflin Company..
- Glaser, B., & Strauss, A. (1967). Grounded theory: The discovery of grounded theory. *Sociology the journal of the British sociological association*, 12, 27-49.
- Grover, S., & Pea, R. (2013). Computational thinking in K-12: A review of the state of the field. *Educational researcher*, 42(1), 38-43.
- Hammond, M. (2010). What is an affordance and can it help us understand the use of ICT in education? *Education and Information Technologies*, 15(3), 205-217.
- Harlow, D. B., Dwyer, H. A., Hansen, A. K., Iveland, A. O., & Franklin, D. M. (2018). Ecological Design-Based Research for Computer Science Education: Affordances and Effectivities for Elementary School Students. *Cognition and instruction*, 1-23.
- Hirose, N. (2011). Affordances, effectivities, and extension of the body. In W. Tschacher & C. Bergomi (Eds.), *The implications of embodiment: Cognition and communication*. (pp. 231-252). Charlottesville, VA: Imprint Academic.
- Hirose, N. (2002). An ecological approach to embodiment and cognition. *Cognitive Systems Research*, 3(3), 289-299.
- Ionescu, T., & Vasc, D. (2014). Embodied cognition: challenges for psychology and education. *Procedia-Social and Behavioral Sciences*, 128, 275-280.
- Johnson, M. (2008). The meaning of the body. In *Developmental perspectives on embodiment and consciousness* (pp. 19-43). Psychology Press.
- Johnson, C. W., & Parry, D. C. (2016). *Fostering social justice through qualitative inquiry: A methodological guide*: Routledge.
- Kopcha, T. J., McGregor, J., Shin, S., Qian, Y., Choi, J., Hill, R., ... & Choi, I. (2017). Developing an integrative STEM curriculum for robotics education through Educational Design Research. *Journal of Formative Design in Learning*, 1(1), 31-44.



- Israel, M., Pearson, J. N., Tapia, T., Wherfel, Q. M., & Reese, G. (2015). Supporting all learners in school-wide computational thinking: A cross-case qualitative analysis. *Computers & Education*, 82, 263-279.
- Lakoff, G., & Núñez, R. E. (2000). Where mathematics comes from: How the embodied mind brings mathematics into being. *AMC*, 10, 12.
- Liben, L. (2008). Representational development and the embodied mind's eye. *Body in mind, mind in body: Developmental perspectives on embodiment and consciousness*, 191-224.
- Liben, L. S. (2012). Embodiment and children's understanding of the real and represented world *Developmental perspectives on embodiment and consciousness* (pp. 207-240): Psychology Press.
- Linderoth, J. (2012). Why gamers don't learn more: An ecological approach to games as learning environments. *Journal of Gaming & Virtual Worlds*, 4(1), 45-62.
- Lye, S. Y., & Koh, J. H. L. (2014). Review on teaching and learning of computational thinking through programming: What is next for K-12?. *Computers in Human Behavior*, 41, 51-61.
- Malafouris, L. (2013). *How things shape the mind*: MIT Press.
- Michaels, C. F., & Palatinus, Z. (2014). A ten commandments for ecological psychology. *Routledge handbook of embodied cognition*, 19-28.
- Richardson, M. J., Shockley, K., Fajen, B. R., Riley, M. A., & Turvey, M. T. (2009). Ecological psychology: Six principles for an embodied-embedded approach to behavior. In *Handbook of Cognitive Science* (pp. 159-187).
- Rietveld, E., & Kiverstein, J. (2014). A rich landscape of affordances. *Ecological Psychology*, 26(4), 325-352.
- Rowlands, M. (2010). *The new science of the mind: From extended mind to embodied phenomenology*: MIT Press.
- Sengupta, P., Kinnebrew, J. S., Basu, S., Biswas, G., & Clark, D. (2013). Integrating computational thinking with K-12 science education using agent-based computation: A theoretical framework. *Education and Information Technologies*, 18(2), 351-380.
- Sullivan, A., & Bers, M. U. (2016). Robotics in the early childhood classroom: learning outcomes from an 8-week robotics curriculum in pre-kindergarten through second grade. *International Journal of Technology and Design Education*, 26(1), 3-20.
- Wilson, M. (2002). Six views of embodied cognition. *Psychonomic bulletin & review*, 9(4), 625-636.
- Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, 49(3), 33-35
- Yadav, A., Mayfield, C., Zhou, N., Hambrusch, S., & Korb, J. T. (2014). Computational thinking in elementary and secondary teacher education. *ACM Transactions on Computing Education (TOCE)*, 14(1), 5.
- Young, M. (2013). An ecological psychology of instructional design: Learning and thinking by perceiving-acting systems *Handbook of research on educational communications and technology* (pp. 180-188): Routledge.
- Yuen, T., Stone, J., Davis, D., Gomez, A., Guillen, A., Tiger, E. P., & Boecking, M. (2015). A model of how children construct knowledge and understanding of engineering design within robotics focused contexts. *International Journal of Research Studies in Educational Technology*, 5(1).

# Bugs as a Nexus for Emergent Peer Collaborations: Contextual and Classroom Supports for Solving Problems in Electronic Textiles

Deborah A. Fields, Utah State University, [deborah.fields@usu.edu](mailto:deborah.fields@usu.edu)  
Gayithri Jayathirtha, University of Pennsylvania, [gayithri@gse.upenn.edu](mailto:gayithri@gse.upenn.edu)  
Yasmin B. Kafai, University of Pennsylvania, [kafai@upenn.edu](mailto:kafai@upenn.edu)

**Abstract:** Few studies have examined the role of failure in more open-ended situations where problems develop as a consequence of designing projects and where collaborations can emerge as an outgrowth of debugging said problems. In this paper, we explore the peer-to-peer collaborations that emerge spontaneously in the context of coding, crafting and design bugs within open-ended design activities, specifically an electronic textiles unit for secondary students taught over 10-12 weeks in introductory computer science classes. Examining observations from three introductory computer science classrooms, we address the following research questions: (1) How and what kinds of peer-to-peer collaborations emerged in unstructured ways, especially around bugs in open-ended projects? and (2) What curricular, spatial, social, and teacher supports allowed these interactions to emerge and flourish? In the discussion, we consider implications for supporting similar types of emergent collaborative learning in open-ended computational making designs.

## Introduction

One could argue that failure in its various forms has more often been seen as an impediment rather than a complement to successful collaboration (Barron, 2003). Indeed, failure can lead to stable patterns of discourse in a classroom community with troubling consequences for students' ideas about themselves (DeLiema, 2017). However, the recent success of "productive failure" (Kapur, 2008) has put failure into a different light: it highlights the counterintuitive notion that failure can precede later success in learning. Yet the research on productive failure has generally emphasized tight structures both in collaborative groups and in the sequence of planned ill- and well-structured problems (e.g., Kapur & Kinzer, 2009, Kapur, 2008). Missing in these studies is an opportunity to examine the role of failure in more open-ended situations where problems develop as a consequence of designing projects and where peer collaborations can emerge as an outgrowth of debugging said problems.

Problems that stem from designing open-ended projects, such as those common in most STEM-oriented maker activities (Peppler, Halverson & Kafai, 2016), occur frequently in many situations where groups are not pre-designated. In open-ended maker activities, temporary failures or unexpected bugs are not just hindrances but also opportunities for learning, when and if students reach outside of their immediate work to a wider group of people and resources (e.g., Sheridan et al., 2014). Reaching out can generate emergent collaborations, that occur quite frequently in makerspaces as students help each other, invent uses for technology, and catch ideas from each other (Halverson, Litts & Gravel, 2018). This raises questions about how peers help each other and what attributes of the broader environment facilitate such improvised collaborations, for even spontaneous collaboration does not take place in a vacuum.

In this paper, we explore the peer-to-peer collaborations that emerge spontaneously in the context of bugs (or problems) within open-ended design activities, specifically an electronic textiles unit for secondary students taught over 10-12 weeks in introductory computer science classes (Fields et al., 2018). Electronic textiles (e-textiles) involve programmable circuits hand-sewn onto soft objects like clothing and stuffed animals, with conductive thread, LEDs, digital sensors, and sewable microcontrollers, providing a space for creating personally relevant computational artifacts (Buechley, Peppler, Eisenberg & Kafai, 2013). Making an e-textile artifact involves learning not only about crafting, circuitry and code but also about identifying, isolating, and fixing bugs at the intersection of these domains (Fields, Searle, & Kafai, 2016; Jayathirtha, Fields & Kafai, 2018). Whereas the study of collaboration has often focused on structural supports that teachers or computer environments provide to ensure success, this study focuses on how peers help each other progress through failure and the qualities of the broader learning environment that support these emergent collaborations. The particular context of our study includes three introductory computer science high school classrooms with 69 students where different teachers at separate public secondary schools led an e-textiles unit in which students created a series of four e-textile projects (Fields et al., 2018). In this paper we explore 1) How and what kinds of peer-to-peer collaborations emerged in unstructured ways (i.e., outside of assigned collaborative groupings), especially around bugs in open-ended projects? 2) What curricular, spatial, social, and teacher supports allowed these interactions to emerge and

flourish? In the discussion we consider implications for supporting similar types of emergent collaborative learning in open-ended computational making.

## Background

Debugging has long been noted as a productive, if not particularly collaborative, site of learning (e.g., Papert, 1980). As a site of problem solving, debugging is recognized as a key computational thinking practice in engineering and computing that is essential but often overlooked in K–12 classrooms (College Board, 2016; McCauley et al., 2008). Debugging computational issues requires a deep and systematic understanding of the program along with the programming language and environment (i.e., McCauley et al., 2008). The ability to debug or troubleshoot may be especially challenging for novice programmers who lack experience in seeing programs as a whole and systematically identifying, testing, and solving problems (Vessey, 1985). Yet while there is significant research around tools and programming environments designed to support learners through the process of debugging (e.g., Ko & Myers, 2004; Sorva, Lönnberg, & Malmi, 2013), there is relatively little research about other kinds of pedagogical, social, and environmental structures that can support learners dealing with various debugging challenges. Understanding learning environments that support debugging is especially important since the complexity associated with debugging demands not just programming skills but also other skills such as decision-making, emotional intelligence, and perseverance (e.g., Patil & Codner, 2007).

Within the field of computer science education, collaboration is generally understudied even if it is recognized as helpful. Most prominent is the research on pair programming where a team of two learners collaborate on one project, designing, coding, testing and debugging on a single machine with students periodically changing roles as “driver” and “navigator” (Williams & Kessler, 2000). While this model illuminates the social aspect of problem solving in programming contexts, this more formal, structured model of collaboration does not include other possible modes of collaboration when more than a pair of learners attempt to support each other. Further, these formal collaboration arrangements become more difficult to maintain when learners are programming in hybrid computing environments such as robotics and e-textiles. For example, the studies about collaboration in making e-textiles or similarly crafted computational artifacts have shown students splitting up their learning in inequitable ways rather than learning productively from each other (e.g., Litts et al., 2017). Moreover, a focus on only formal ways of pairing up students misses opportunities to explore other more open-ended forms of collaboration and emergent social supports where students can draw help from peers other than their immediate partners.

In this study then, we focus on the nature and contexts of emergent peer learning while debugging in a classroom environment, how it emerged (because of the teacher, peers, or through other means), and how it mattered to individual students. This type of peer-to-peer collaboration has received much less attention in research because most studies have focused on organized small groups, with students taking on various roles or with some students designated as more experienced experts or less experienced novices in situations intended to generate “peer pedagogy” where peers educate each other (e.g., Ching & Kafai, 2008). Yet teachers are already developing environments where more emergent peer pedagogy, not structured in specific groups, can take place. For instance, a prior study on the e-textiles curricular unit for introductory computing classes documented productive teacher practices that supported equity, namely by legitimizing peer expertise and supporting iterative learning (Fields et al., 2018). The researchers observed that teachers modeled their own and their students’ mistakes to the whole classroom, lowering the risk associated with failure and allowing students to provide ideas to solve problems in the projects. Sometimes, especially in situations where the teacher could not give individualized help to each student, the teacher would train one student in a task (e.g., a crafting or computing technique) with the expectation that that student would then share that knowledge with others. These and other practices, structures, and environmental features may support productive peer pedagogy.

The many peer-to-peer learning collaborations that emerged beyond formal partnerships in our study appeared to be highly significant to students’ continued perseverance in the e-textile unit. So, we sought to understand how these specific collaborations emerged, what relevance they had to individual students, and what aspects of the larger classroom environment facilitated these types of interactions.

## Methods

### Context and participants

Over the course of two years, we developed an e-textiles unit for the *Exploring Computer Science* (ECS) curriculum, a year-long introductory computing course with equity-focused and inquiry-based teaching (Goode, Chapman, & Margolis, 2012). This paper focuses on the second year of this e-textiles unit implementation, where three teachers led the unit separately in each of their respective schools. The e-textiles unit took place over 10–12



weeks and consisted of a series of four projects: 1) a paper-card using a simple circuit, 2) a wristband with three LEDs in parallel, 3) a classroom-wide mural project completed in pairs that incorporated two switches to computationally create four lighting patterns (the only collaborative project), and 4) a “human sensor” project that used handmade sensors to create conditions for lighting effects (see Kafai & Fields, 2018). Each project allowed increasing flexibility in design and personalization in the context of learning challenging new computing concepts. The e-textile projects provided many opportunities for debugging since problems occurred in the code, the circuitry, and the physical designs, requiring students to test and isolate problems, often fixing multiple co-occurring issues that added to the complexity and challenge of the projects (Kafai et al., 2014; Fields et al., 2016; Jayathirtha et al., 2018). The unit also drew on ECS practices such as pair programming during certain coding lessons.

The three schools were located in a large metropolitan school district and served a large number of students who came from ethnic/racial groups that are traditionally underrepresented in CS. At the three schools, 54-95% of students were identified as socioeconomically disadvantaged as defined by the state, and were ethnically diverse (40-90% Hispanic or Latino/a, 1-25% White, 4-43% African American, 0-18% Asian, 0-10% Filipino). The classes each included 20-35 students from 9th-12th grade (14-18 years old) and were diverse in terms of and gender ratios (36-66% girls). All teachers received six days of professional development over two years: three days in the first year of the curriculum development focused on design and content learning, and three days in the second year focused on supportive pedagogical practices.

## Data collection and analysis

Data for our analysis was drawn from fieldnotes, student artifacts, and interviews across the three classrooms. Two researchers, including the first author who was also a lead designer of the unit, collected weekly field notes in each classroom adding up to 39 fieldnotes. In addition to observational fieldnotes, we collected students’ design notebooks, journals, and end-of-unit portfolios where students reflected on challenges and learning during the unit. Further, we interviewed 12 students from each classroom in focus groups at the end of the unit and interviewed teachers before, during, and after the unit. Of note, the original focus our data collection was on students’ learning of computational and circuitry concepts, which matched our research agenda. However, we recognized the central role of unplanned peer-to-peer collaborations in students’ learning and perseverance in the unit after initial analysis of students’ progressions of learning with e-textiles. This led us to analyze emergent collaboration in its own rite.

In analyzing the data, we first sought to identify all examples of emergent peer-to-peer collaboration across our observations. We defined this to include only student collaborations outside of teacher-directed groups (i.e., pair programming lessons) or formal partnerships (i.e., the collaborative pair-work in project #3) that related to student work (i.e., excluding casual banter unrelated to classwork). Drawing from the principles of grounded theory (Charmaz, 2014), two researchers initially coded a set of three field notes (close to 10% of the total), to come to agreement on applications of this definition, and coded a second set to verify consistency. Then one researcher culled all examples of emergent peer-to-peer collaboration in the entire dataset. We then analyzed these moments holistically (Miles, Huberman & Saldana, 2014), often returning to fieldnotes or related data (i.e., portfolios where students reflected on problems), to understand what stimulated those moments of emergent collaboration, the larger context that surrounded them, and their implications on the students. Next, using a two-step open coding analysis (i.e., developing initial codes, analyzing the dataset in full, re-visiting and editing the codes, and re-analyzing the dataset in full), we jointly developed the following categories which we will elaborate on in the findings section:

- *Stimuli for collaboration*: Bugs, verbal exclamations, help other than debugging, stimulus from a lesson, other.
- *Subject of collaboration*: Problem-solving, feelings of camaraderie, extra pair of hands, other.
- *Enabling environmental factors*: Physical proximity, overseeing/overhearing, curricular affordances, teacher practices, distributed expertise, adult brokerage, earlier friendship

Overall, our analysis was similar to Sawyer’s work (2012) in that there was a focus on individual actions that was used to identify the events of emergent collaboration, which then guided a deeper sociocultural analysis of the contextual factors around these events.

## Findings

We identified five types of stimuli for emergent peer collaboration: bugs (a problem in a project or task); exclamations (students saying something loudly, such as something that went well); general help (including

learning a new technique, getting advise on the aesthetics or appearance of a project or completing a necessary task); following through on a lesson-based task that was not intended for group work, or a general “other”. Figure 1 shows the frequency of instances in each of these areas as well as the type of collaboration that occurred: solving a problem, building camaraderie between peers, working together not on a problem but on something that required “extra hands” (i.e., holding something), and of course, “other.”

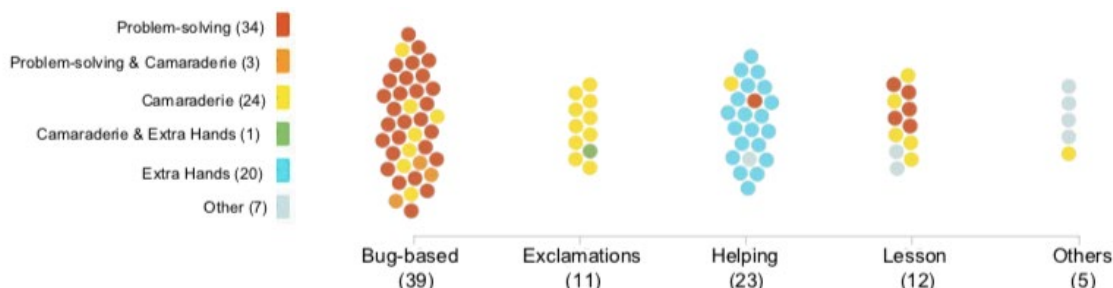


Figure 1. Graphical visualization of all the instances of emergent collaborations across the themes.

Since the majority of the instances that involved bugs (a little less than half of all emergent collaborations), we focus the remainder of this paper on investigating this area more deeply. Bug-related events might involve realizing a breakdown, identifying a bug, fixing, testing or otherwise debugging a project or celebrating fixing a bug. Collaborations stimulated by bugs included directly helping in problem-solving (red and orange dots in Figure 1) or providing emotional support (i.e., “camaraderie”—orange and yellow) by fostering a sense of camaraderie around making mistakes by encouraging, being comical or even heckling. In order to share the wide variety of technical and emotional supports that students provided to each other, below we explicate three examples of peer collaboration which illuminate the diversity of supports and immediate contextual factors that enabled the collaboration. We also include information about the significance of these moments for students who were involved in these collaborations. At the end we consider the broader contextual factors that enabled these emergent collaborations, including the tools, spaces, classroom and lesson structures, classroom practices, and student and teacher roles.

## Observations of emergent collaborations

### Initiating peer-to-peer debugging

One way that bugs became a source of collaboration was when students noticed another student’s bug and inserted themselves into the debugging process without being invited. For instance, this happened with Andrea, a student who had earlier been identified by her teacher as a ‘middle of the class’ student who struggled with motivation in the class. Unusually for her (teacher post-interview, May 24, 2017), Andrea was one of the first to finish her final project (the human sensor project) and noticed that a friend sitting at the same table, Maria, was struggling with her project. The collaboration proceeded as follows:

First, Andrea noticed that Maria was having trouble with her project. Andrea shifted her seat to be next to Maria and began to interact with Maria’s project. When uploading the code, an error message about the “port” came up, and the girls realized that the USB cable was not working. They switched out the cable and were able to upload the code, but no lighting patterns were triggered when the sensor (two aluminum foil patches) was activated by their touch. Andrea set her computer next to Maria’s to check the code. Andrea went line by line through the code with Maria to see if was correct. ‘I think it’s the 1000 [sensor range]. Mine is set to 900 and it works better; maybe your range is too small,’ Andrea said. In this, Andrea identified that the sensor threshold was too high, and the ranges within the conditional statements did not trigger the outputs Maria had programmed. Once they changed the ranges, two of the four patterns in Maria’s project worked. Maria then began to debug her own code, identifying some places in her lighting patterns that were not giving the effects she intended. Andrea continued to sit next to Maria for the rest of the time, playing with a fidget spinner and answering questions whenever Maria asked for help (field note, May 18, 2017).

This spontaneous collaboration began with a bug in Maria’s project apparent without any requests for assistance. In fact, spotting a simple USB cord problem led to a series of bugs including the more challenging customization of sensor ranges within a set of conditional statements. Notably, both the teacher and researchers noticed a change in Andrea during the e-textiles unit, moving from hesitation about computing and coding to being “way more proud and way more into the work” (teacher post-interview, May 24, 2017). This change was

visible in Andrea's confident approach to her friend and tablemate Maria, and in the ways that she both helped identify and fix problems and stood by while Maria continued to debug on her own.

### Soliciting help on peer-to-peer debugging

Another type of emergent peer collaboration involved students explicitly soliciting another's help to debug their projects. Many a times students requested help from peers they considered more expert in a specific area, from threading a needle to checking a circuit diagram to fixing code. This recognition of expertise may have come from a teacher or researcher suggesting getting help from a particular student or, more often, from prior knowledge of a student's expertise from earlier in the class, as it did with Jesse below.

During the fourth project, Jesse had successfully sewn and tested his lights and had sewn his sensors patches on but was "terrified" to start the coding of his sensors, which involved reading the sensors and deciding on a set of ranges that would work with different levels of squeezing the two patches (fieldnote, Mar. 07, 2017). To start on the coding, he got up from his seat and went to the other side of the classroom to seek help from Gencio, a student who had already worked through much of his sensor code (with the help of a researcher) and who was known as a very proficient coder in the class. After getting assistance from Gencio, Jesse came back to his table and almost immediately switched roles from being helped to helping, looking at his neighbor Joyce's code and pointing out an incorrectly (or inconsistently) named variable. Joyce quickly corrected her error, but not before the interaction drew the attention of Felix, sitting at the same table, who teased Joyce for not listening carefully.

The whole sequence of events demonstrates an example of cascading help that was provided from one person to another. First, the researcher helped Gencio and his partner with coding and debugging the sensors on their projects. Later, Jesse approached Gencio for help. Then Jesse pointed out an error to Joyce, and the whole table became involved in Felix's teasing about Joyce's coding bug. Conversations such as these where students developed a sense of camaraderie by heckling, joking and even encouraging each other were observed frequently as demonstrated in the Dante and Juan's story (see next section).

### Peer camaraderie over failure

One other type of collaboration that occurred around debugging included many friendly conversations where peers sympathized about bugs. Often prompted by a student's encounter with some mistake or issue, these conversations became places where students not only shared their individual problems, but also debated their causes and considered possible solutions. Along with promoting a sense of camaraderie and/or sympathy, these moments provided much needed emotional support for one to persevere through moments of frustration and sense of failure. For example, below is a fieldnote record (Jan. 26, 2017) illustrating Dante's experience when debugging his dysfunctional wristband (project #2) led to a spontaneous conversation, inviting his nearby classmate's attention. During this episode, Dante shifted from discouragement to sharing laughter with peers about common mistakes, highlighting the way camaraderie influenced his participation.

Researcher: How's it going?

Dante: I can't sew. Yeah, I keep messing up like 30 times. *(Other students giggle.)*

Researcher: What did you mess up on? *(Others are still snickering.)* Everybody else knows what you're messing up on?

Dante: It's 'cause I didn't pull the string all the way, so I had loops in the back...

Juan: Yeah, but his first mistake was getting that color, though.

Dante: Oh yeah. Yeah, my first mistake was getting this [gray] color felt, so I barely see the [gray conductive] string.

Researcher: Ohhh, it's hard to see!... So that was the first problem, the color?

Dante: Yeah, because I didn't know what I was doing so I was just pulling the strings, and before I knew it, I was pulling apart... *(Points to frayed areas of felt on his project.)*

Researcher: Oh, 'cuz you can't tell what's the string and what's the fabric.... *(Juan laughs)* And Juan, why're you laughin' at him?

Juan: 'Cuz I made the same mistake!

*(Juan, sitting across the table from Dante, starts to describe how messy the back of his wristband was, after he thought he did a good job stitching and pulling the thread all the way through. He talks on for a minute.)*

Juan: Yeah, and when you look at the back... (*Several others laugh, including Dante*).

Dante's visible errors—large knotty clumps of thread with fabric the same grey color of the thread—invited attention and laughter from his classmates, providing an opportunity for Juan to share his own similar mistakes. While the laughter of Dante's classmates may initially appear off-putting, the tone of the conversation implies that the laughter was more in sympathy than in jest, effectively lessening the significance of a single mistake. In fact, Dante joined in on the laughter at the end, sharing a sense of camaraderie over common problems. Although Dante never successfully completed his wristband, this did not prevent Dante from finishing his next two (more difficult) projects. In the end, even creating a fully-functional final project which involved a spatially complex, three-dimensional interactive ball created attaching four pieces of felt.

## Designs and supports for emergent collaboration

Analyzing the occurrences of emergent collaboration around bugs (including those above as well as our entire dataset), here we consider the features of the learning environment that supported this phenomenon. Overseeing, overhearing, and close proximity to peers were important features that related to physical aspects of both the projects and the spaces in classrooms. For instance, bugs stood out visibly in the physical *tangibility* of e-textiles. Mistakes such as knotty threads or malfunctioning lights were visible to others within eyesight, effectively serving as “open tools” (Hutchins, 1995) that triggered collaboration and, as seen above with Dante and Juan, commiseration over problems. Further, all three classrooms had *open physical spaces* where students were grouped at tables and could see each other's work. A number of collaborations happened between students sitting next to each other or across the table from each other (e.g., Andrea and Maria). Muttered frustrations or exclamations also led to conversations and collaborations about problems, even when problems were not physically visible.

More subtle than visible and audible awareness of bugs, the *curricular structures* of the e-textile unit further enabled emergent collaborations. The design constraints for the projects meant that each student had to learn similar concepts (i.e., coding conditionals for human touch sensor patches in project #4) while allowing for personal customization (Kafai & Fields, 2018). This combination of constraints and personalization meant that students could legitimately learn from each other and even copy techniques of code, circuitry, or craft while still making something unique. For instance, looking at a peer's code served as a stepping stone for students like Maria or Jesse to debug their projects. Further, the curriculum suggested regular formal collaborations in certain lessons, including pair programming, a collaborative project (project #3), and whole class discussions. This provided precedence and practice for students to work together, and many students either returned to earlier partnerships or used knowledge about peers' expertise gained from class discussions to seek help.

The social structures promoted in the curriculum combined with prior *social networks* in the classes to facilitate emergent collaborations. Adding to the formal collaborative pairings and discussions, students used their own prior networks of friendship in providing help, seeking assistance, and sympathizing over problems. Many emergent collaborations took place between students who had been long-time friends, as was the case of Andrea and Maria. In addition, some students like Dante stated that they developed new friendships during the unit more than at any other time of the school year, expanding social networks.

Finally, *teachers' practices* and *classroom management* played both indirect and direct roles in emergent collaborations around bugs. For instance, unspoken classroom rules allowed for the *mobility* of students and for *informal conversation* about problems. Though students were expected to be on task, this did not mean that they had to sit silently at their own seat as they might be expected to in other school contexts. In addition, teachers often intervened more directly to support peer-to-peer learning by brokering help between students. Further, as we found in our study of teacher practices in the unit (Fields et al., 2018), teachers frequently legitimized students' diverse and developing expertise, highlighting students' knowledge from their earlier mistakes in classroom reflections and sharing of “tips and tricks”. This supported the idea of turning to students, not just teachers, as sources of knowledge. Open physical spaces, tangible artifacts, project constraints, formal collaborations, informal social networks, combined with classroom practices to facilitate emergent peer collaborations around bugs.

## Discussion

In this paper we described unplanned peer-to-peer collaborations that emerged around problems, or bugs, in students' open-ended e-textile designs. Though others have pointed out the collaborations that happen around idea-generation and technique diffusion in makerspaces (Halverson et al., 2018), in our study problem-naming, fixing and helping were a nexus for emergent peer collaboration. Student-to-student collaborations around bugs

resulted in just-in-time assistance with troubleshooting, encouragement, and even humor that helped students' progress and persevere through problems. In doing so students took up roles of leadership and expertise while developing friendships in the class and inserting humor into an often-frustrating design process. Unlike the engineered moments of "productive failure" scenarios (e.g., Kapur, 2008), there is an ample supply of bugs in e-textile designs that arise without structured planning. Together these findings not only refine our understanding of student collaborative experiences in the face of failures but also highlight some of the key contextual factors that turn these moments into productive ones.

In discussing various environmental features that support emergent collaboration, we considered the roles of physical, curricular, social, and teacher attributes. Tangible artifacts and open spaces combined with social norms that allowed movement, conversation, and improvisation afforded opportunities for students to see, hear, and converse with each other over challenges. In settings where making or constructing are the focus and where the teacher is not the only expert in the room, it is important to highlight the more tacit qualities of learning environments and classroom practices that facilitate learning. We do not assume to have captured all environmental aspects that support emergent peer collaboration. Rather we hope that this study provides a starting point for exploring bug-focused collaborations and contextual factors that support emergent collaborations. Much more awaits further study, such as potential relationships between types of problems and the collaborations that emerge around them, the role of different tools and representations in the problem space in inviting collaboration (or not), and further environmental supports (structures, spaces, norms, practices) that facilitate or constrain peer collaboration. More work is also needed on the social-emotional qualities of emergent collaborations, for instance the "camaraderie" identified in this study, which played an important role in making "failure" and bugs more socially acceptable to students like Dante and Juan, facilitating not just problem-solving but emotional robustness and overall perseverance during the course of the unit.

Much work in computer supported collaborative learning has focused on designing direct scaffolds for participation, whether in the form of group roles, group structures, on-screen or in-class scaffolds, or tasks that facilitate collaboration. One of the challenges with such fixed arrangements is that they provide little agency for students to develop their own competencies in dealing with frequent unforeseen challenges that occur while making computational artifacts outside the structured settings. Emergent peer collaboration provides a means to acknowledge and support student agency, expertise, and creativity. Designing environments for emergent collaboration, including in relation to more formal collaborations, may open up even greater opportunities for student learning and development.

## References

- Barron, B. (2003). When Smart Groups Fail. *Journal of the Learning Sciences*, 12(3), 307-359.
- Buechley, L., Peppler, K., Eisenberg, M., & Yasmin, K. (2013). *Textile Messages: Dispatches from the World of E-Textiles and Education. New Literacies and Digital Epistemologies. Volume 62*. Peter Lang Publishing Group. 29 Broadway 18th Floor, New York, NY 10006.
- Charmaz, K. (2014). *Constructing grounded theory*. Sage publications, Thousand Oaks, CA.
- Ching, C. C., & Kafai, Y. B. (2008). Peer pedagogy: Student collaboration and reflection in a learning-through-design project. *Teachers College Record*, 110(12), 2601-2632.
- College Board (2016). *AP Computer Science Principles: Course and Exam Description Effective Fall 2016*. College Board: New York, NY. Retrieved from: <https://secure-media.collegeboard.org/digitalServices/pdf/ap/ap-computer-science-principles-course-and-exam-description.pdf>.
- DeLiema, D. (2017). Co-constructed failure narratives in mathematics tutoring. *Instructional Science*, 45(6), 709-735.
- Fields, D. A., Kafai, Y. B., Nakajima, T. M., Goode, J., & Margolis, J. (2018). Putting making into high school computer science classrooms: Promoting equity in teaching and learning with electronic textiles in *Exploring Computer Science. Equity, Excellence, and Education*, 51(1), 21-35.
- Fields, D. A., Searle, K. A., & Kafai, Y. B. (2016). Deconstruction kits for learning: Students' collaborative debugging of electronic textile designs. In *FabLearn '16, Proceedings of the 6th Annual Conference on Creativity and Fabrication in Education*, ACM, New York, NY, 82-85.
- Goode, J., Chapman, G., & Margolis, J. (2012). Beyond Curriculum: The Exploring Computer Science program. *ACM Inroads*, 3(2), 47-53.
- Halverson, E., Litts, B., & Gravel, B. (2018). Forms of Emergent Collaboration in Maker-based learning. In *Proceedings of the International Conference of the Learning Sciences*, 2018, 921-924.
- Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT.

- Jayathirtha, G., Fields, D., & Kafai, Y. (2018). Computational Concepts, Practices and Collaboration in High School Students' Debugging Electronic Textiles projects. *Proceedings of the International Conference on Computational Thinking Education 2018*. Hong Kong: The Education University of Hong Kong. 27-32.
- Kafai, Y., Fields, D. (2018). Some Reflections on Designing Constructionist Activities for Classrooms. In *Proceedings of the Constructionism Conference 2018*. Vilnius, Lithuania. 601-607.
- Kafai, Y., Fields, D., & Searle, K. (2014). Electronic textiles as disruptive designs: Supporting and challenging maker activities in schools. *Harvard Educational Review*, 84(4), 532-556.
- Kapur, M. (2008). Productive Failure. *Cognition and Instruction*, 26,(3), 379-424.
- Kapur, M., & Kinzer, C. K. (2009). Productive failure in CSCL groups. *International Journal of Computer-Supported Collaborative Learning*, 4(1), 21-46.
- Ko, A. J., & Myers, B. A. (2004, April). Designing the whyline: a debugging interface for asking questions about program behavior. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 151-158. ACM.
- Litts, B. K., Lui, D. A., Widman, S. A., Walker, J. T., & Kafai, Y. B. (2017). Reflections on Pair E-Crafting: High School Students' Approaches to Collaboration in Electronic Textiles Projects. In *Proceedings of the International Conference of Computer Supported Collaborative Learning 2017*, Philadelphia, PA (2) 569-572.
- Miles, M. B., Huberman, A. M., & Saldana, J. (2014). *Qualitative data analysis: A methods sourcebook*. Sage publications, Thousand Oaks, CA.
- McCauley, R., Fitzgerald, S., Lewandowski, G., Murphy, L., Simon, B., Thomas, L., & Zander, C. (2008). Debugging: a review of the literature from an educational perspective. *Computer Science Education*, 18(2), 67-92.
- Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas*. Basic Books, Inc..
- Patil, A., & Codner, G. (2007). Accreditation of engineering education: review, observations and proposal for global accreditation. *European journal of engineering education*, 32(6), 639-651.
- Peppler, K., Halverson, E., & Kafai, Y. B. (Eds.). (2016). *Makeology: Makerspaces as learning environments* (Vol. 1). Routledge.
- Sawyer, K. (2012). Extending sociocultural theory to group creativity. *Vocations and Learning*, 5(1), 59-75.
- Sheridan, K., Halverson, E. R., Litts, B., Brahms, L., Jacobs-Priebe, L., & Owens, T. (2014). Learning in the making: A comparative case study of three makerspaces. *Harvard Educational Review*, 84(4), 505-531.
- Sorva, J., Lönnberg, J., & Malmi, L. (2013). Students' ways of experiencing visual program simulation. *Computer Science Education*, 23(3), 207-238.
- Vessey, I. (1985, December). Expertise in Debugging Computer Programs: Situation-Based versus Model-Based Problem Solving. In *International Conference of Information Systems 1985*, 18, 288-302.
- Williams, L. A., & Kessler, R. R. (2000). All I really need to know about pair programming I learned in kindergarten. *Communications of the ACM*, 43(5), 108-114.

## Acknowledgments

This work was supported by a grant from the National Science Foundation to Yasmin Kafai and Michael Eisenberg (#1742140). Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of NSF, the University of Pennsylvania, or Utah State University. Special thanks to Tomoko Nakajima for her help with data collection and to Alecia Magnifico, Debora Lui, Justice T. Walker, and Mia Shaw for their valuable feedback.

# Supporting Collaborative Curriculum Customizations Using the Knowledge Integration Framework

Korah J. Wiley, Allison Bradford, and Marcia C. Linn  
korah.wiley@berkeley.edu, allison\_bradford@berkeley.edu, mclinn@berkeley.edu  
University of California, Berkeley

**Abstract:** Teachers are currently facing a major instructional challenge, namely, supporting students to meet the three dimensional learning goals of the Next Generation Science Standards without adequate curriculum materials to do so. In this paper, we report the design and outcomes of professional development activities that support teachers and researchers to collaboratively customize Web-based Inquiry Science Environment (WISE) units to help students develop coherent science knowledge. The WISE units and professional development activities were developed using the Knowledge Integration (KI) Framework. We show that the KI framework functioned as an effective scaffold to support teachers in modifying their teaching practice to make curriculum customizations that are evidence-based and aligned with a theory of learning. We discuss how our study results informed the design of an online curriculum customization and implementation interface and offer the principles of the KI framework as design principles for the development of other collaborative professional development endeavors.

**Keywords:** curriculum customization, professional development workshop, teacher-researcher collaborative learning, NGSS, knowledge integration

## Introduction

We report on the design and impact of a professional development workshop that positions teachers and researchers as collaborative partners in customizing Web-based Inquiry Science Environment (WISE) units to align with the Next Generation Science Standards (NGSS, 2013). To the workshop, teachers brought their teaching expertise and used their experience implementing the units and logged student data to improve the unit. Researchers brought their expertise in designing the units following the Knowledge Integration (KI) pedagogical framework and used the framework to guide the professional development activities. We hypothesized that, if aligned with curriculum activities, the four steps of the KI process (i.e. eliciting ideas, adding ideas, distinguishing ideas, and reflecting on ideas) would serve as instructional strategies and scaffold teachers' application of the underlying constructivist theory of learning. We further hypothesized that, if aligned with the workshop activities, the four KI steps would give teachers and researchers an experiential knowledge of the process, facilitate the integration of their respective knowledge, and foster the collaborative customization of WISE units for NGSS-alignment and variegated classroom implementation. We discuss how the results from this study informed the design of an online interface on the WISE platform to support teachers to customize and implement NGSS-aligned curricula in ways that helps students develop coherent, three-dimensional science knowledge.

## Background

Since its publication in 2013, the NGSS (NGSS, 2013) has been widely adopted by U.S. states, often without accompanying curriculum materials. During initial implementation of WISE units, many teachers reported interest in better aligning their instruction with NGSS, motivating this study. Given the new, three-dimensional learning goals of the NGSS, developing curricula de novo that can support students to meet these ambitious standards is a formidable task, even for the most adept teacher. In their instructional comparison study, Penuel and Gallagher (2009) showed that when teachers assume the role of curriculum customizer rather than designer, they produce higher quality curriculum, as measured by its ability to support inquiry and promote student science understanding. Building on these findings, we designed a knowledge integration workshop to build teachers' capacity to customize existing curricula rather than design new curricula.

A review of professional development supporting teachers to use technology-enhanced science inquiry curricula found a significant effect for programs that followed the KI framework (Gerard et al, 2011). Successful professional development efforts elicited teachers' ideas, added new ideas to their existing repertoire, and encouraged teachers to distinguish among their ideas, reflect upon their experiences, and develop an



integrated, coherent view of instruction (Linn & Eylon, 2011). The KI framework draws on extensive research supporting a socio-constructivist view of learning.

A key tenet of socio-constructivist learning and knowledge integration is sustained collegial collaboration. In successful professional development efforts, this collaboration involves teachers from multiple schools along with university mentors (Penuel et al., 2007). As teachers collaborate with other education community members in activities such as customization, they negotiate meaning (Lave, 1996), which is a critical aspect of learning.

## Research questions

Research suggests the value of focusing on the customization of existing curriculum materials. However, supporting teachers to incorporate the NGSS into their teaching practices calls for new models of professional development. The KI framework as a professional development model offers promise, especially for the customization of WISE units, since they were designed using the KI framework. Thus, we investigated the following research questions:

1. Can the KI framework guide the design of both curriculum customization and professional development?
2. Which professional development activities:
  - Enable teachers to use pedagogical principles to align existing curriculum materials with NGSS?
  - Support teachers and researchers to collaborate to develop curriculum that enables students to meet the three-dimensional learning goals of the NGSS?

## Methods and materials

We designed and tested professional development activities consisting of a 1.5 day workshop to review and customize WISE units, in-class support during implementation of the customized unit, and post-implementation interviews. The professional development activities followed the KI pedagogy of eliciting ideas about teaching the unit, adding new ideas to customize the unit, distinguishing among ideas during implementation of the customized unit, and reflecting on the experience.

## Participants

The workshop participants included 21 middle school science teachers and 15 researchers. The teachers came from 8 schools across 6 districts and the researchers came from 2 universities. All teachers had implemented at least one WISE curriculum unit prior to attending the workshop.

## WISE curriculum units

WISE units are developed using the KI Framework. They elicit student ideas from their own experiences; engage students in gathering new ideas using embedded models and simulations and hands-on activities; encourage students to distinguish among these ideas by building models, testing alternative views, or critiquing ideas of others; and request reflections in reports or presentations. The WISE units used by teachers in this study are: Photosynthesis, Global Climate Change, Plate Tectonics, and Self-Propelled Vehicles. Each unit features embedded assessments, logs student responses, and captures student activities (e.g. click-stream data). Teachers and researchers can access all logged, de-identified student data via the data export interface in the Grading Tool.

## Customization materials and activities

Before the workshop, in preparation for the customization process, we analyzed the WISE units to identify the NGSS performance expectations addressed in each. Then, we restructured the units into lesson series such that each lesson in the unit targeted a single NGSS performance expectation. To support coherent knowledge building for the target performance expectation, we also structured each lesson in the unit to engage students in each step of the KI process (i.e. one lesson corresponds to engagement in a complete KI cycle). The lessons were designed such that they could be taught in sequence as a unit or as independent lessons.

In addition, we created a diagrammed version of each unit consisting of a 5" x 7" notecard for each lesson with each activity in the lesson briefly described on 3" x 3" sticky notes, color-coded by each KI step (pink: Elicit Ideas; orange: Add Ideas; green: Distinguish Ideas; blue: Reflect On/Revise Ideas, see Figure 2C). On the 5" x 7" notecards was the following lesson information: unit title, lesson title, recommended grade



levels, targeted performance expectation, and a brief description of the learning goals. These tangible, diagrammed versions of the WISE units were the objects of customization and were designed to pilot a prototype version of a unit customization interface on the WISE platform.

After the workshop, the researchers incorporated the customizations of the diagrammed version of each unit into a complete digital version for subsequent post-workshop classroom implementation.

## Data sources and data analysis

To address our research questions, we gathered data corresponding to each of the KI-aligned professional development activities. During the 1.5 day workshop, we documented the customized units for comparison to their previously implemented counterparts. Throughout the workshop, the researchers captured audio recordings and photographs of the workshop activities, and collected teachers' written responses to the following prompts:

- What are some things you learned or have taken away from engaging in this customization process and sharing with other teachers?
- Was this customization process and reflecting on the KI cycle helpful for you in thinking about how to achieve your NGSS and other curricular goals?
- Do you think you could use this customization process for another unit you'd like to run?
- Please share any other reflections or feedback you have from the workshop.

To analyze the customization of the WISE units, we counted the number and type of interleaved KI-aligned, non-WISE activities that teachers added to the diagrammed units. Our analysis of teachers' written reflection consisted of identifying themes related to the customization and collaboration process. For this paper, we analyze the implementation of the customized WISE Photosynthesis unit as carried out by two, 7<sup>th</sup> grade teachers, Mr. Vega and Mr. Harrison. We took field notes during classroom observations and conducted post-implementation interviews to discuss their customization decisions and their overall implementation experiences. We evaluated our observation notes and interview data in terms of their implementation of the workshop customizations and the type of customizations they made during implementation.

## Results

### Knowledge-integrating and experiential activities

The activities of the 1.5 day workshop aligned with the four steps of the KI process (see Table 1) and positioned teachers as experts on their teaching practice and researchers as experts on their curriculum designs. To begin the workshop, the researchers facilitated a whole group discussion to elicit ideas about the function of each dimension of the NGSS performance expectations in terms of lesson development (e.g. disciplinary core ideas provide the lesson content). During this discussion, the researchers introduced the KI framework and invited teachers to share activities they used or could KI process. This activity was designed to highlight the similarities between how the teachers and researchers conceptualize and support the knowledge building process. Each group activity of the workshop (whole group and small group) was designed to highlight the expertise of both teachers and researchers and to create opportunities for expertise sharing.

Table 1: A summary of the professional development activities, their alignment with the KI process, and sample teacher reflections.

| Professional Development Activity        | Knowledge Integration Alignment | Sample Teacher Reflections   |
|--|---------------------------------|--|
| KI Framework for NGSS Lesson Development | Eliciting & Adding Ideas        | <i>"I enjoyed looking at the KI. It is very similar to how we plan our lessons and it was cool to see we were aligned in our thinking."</i>  |
| Identifying KI-supporting Activities     | Eliciting & Adding Ideas        | <i>"[KI] allowed [me] to better think about the learning process. Breaking it down into components and deciding what kind of activities and scaffolding will go into that part of the project"</i> |
| WISE Unit Analysis Using                 | Adding & Distinguishing         | <i>"This is something I will use in the future with other lessons"</i>   |

|   |                                      |  |
|---|--------------------------------------|--|
| Student Data  | Ideas                                | <i>to make sure it is really meeting all areas[, if] students [are] really understanding a lesson.”</i>  |
| Interleaving KI-supporting Activities into WISE Units (Customizing) | Adding & Distinguishing Ideas        | <i>“The customization process and reflecting on the KI cycle are definitely helpful for me. I understand better how each unit is built and how I can interact with it. I can pull different components together to fit the needs of the curriculum or the time constraints”</i>  |
| Reviewing Unit Customizations                                       | Distinguishing & Reflecting On Ideas | <i>“It was really interesting to hear other teachers' reflections on their units. It gave me the chance to think about how I might want to use some similar ideas in the units I plan to run in my classroom.”</i>   |
| Actualizing Customizations Using the WISE Platform                  | Adding Ideas                         | <i>“I added new labs to the WISE project, learning how to use the authoring tool was great to be able to add our new ideas into the project.”</i>  |
| Implementation of Customized Unit                                   | Distinguishing & Reflecting On Ideas | <i>“I deployed instruction in similar fashion to the Knowledge Integration process through a unit designed and planned by [another curriculum provided] for 6th graders learning from instruction that is 3-D and that conforms to the NGSS. I generally follow the same process throughout each unit of instruction.”</i>   |
| Post-Implementation Interviews                                      | Reflecting On Ideas                  | <i>“The [KI] labels on each step [of the customized WISE unit] were useful to me in identifying the stage of learning the students were in; to which, I targeted my instruction or interactions with small groups. I loved how the unit was divided into distinct but incremental lesson[s] for that exact reason, they demarcated conceptual bubbles that can exist on their own but are still a part of something larger, the bigger picture, the overarching idea.”</i> |

The sharing and integration of expertise was evident in the customization activity. For the unit customization activity, teachers were assigned to a small group corresponding to a unit they recently implemented and with researchers who had developed the units. Teachers were asked to think of and write on colored 3 x 3 sticky notes as many activities as they could think of for each step of the KI process (Pink: elicit ideas; Orange: add ideas; Green: distinguish ideas; Blue: reflect on/revise ideas). Teachers were invited to share their KI activity ideas with other workshop participants by placing the sticky notes on a long table centrally located in the meeting room (see Figure 1). The goal of this activity was to further elicit and add ideas regarding topic-specific activities that were accessible to teachers and would support the KI process.

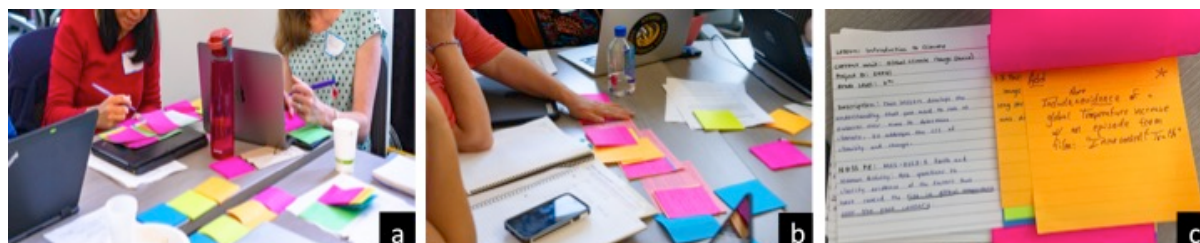


**Figure 1.** Shared interleaved teacher activities.

To further add ideas regarding unit customization, the researchers provided teachers with evidence of their students' learning. Specifically, researchers gave the teachers a random sampling of 30 of their students' responses to a post test assessment item that targeted at least one dimension of an NGSS performance expectation addressed in the unit. Teachers were asked to evaluate the responses to identify areas of strength and weakness in their students' understanding.

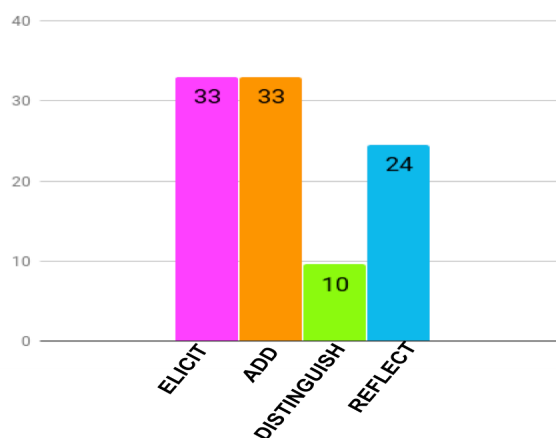
The researchers then invited the teachers to customize the unit by interleaving the activities they previously wrote on the color-coded sticky notes with those on the lesson notecards. The goal of this activity

was to customize the unit in ways that would better support students in developing an integrated understanding of the targeted NGSS performance expectation, using evidence of previous student learning as a point of reference. Teachers were given full license to eliminate, substitute, or add any KI-supporting activity by rearranging the sticky notes on the lesson notecards. In this way, teachers could distinguish their ideas about how to customize the unit in ways that would support their students in the KI process and function within their classroom constraints and resources (see Figure 2A, 2B).



**Figure 2.** WISE unit customization process: Identifying non-WISE, KI-aligned activities (a); Identifying relevant activities to customize the unit (b); Integrating their activities into the diagrammed WISE unit (c).

Teachers worked collaboratively within and across their small groups to exchange ideas and activities to customize their unit. Of all the customizations that teachers made across all the units, 33% (31/94) would elicit students' ideas, 33% (31/94) would add to students' ideas, 10% (9/94) would support students in distinguishing their ideas, and 24% (23/94) would help students reflect on or revise their ideas (see Figure 3).



**Figure 3.** Percentage of unit customizations aligned to each Knowledge Integration step.

In another group workshop activity, teachers were asked to evaluate whether their customized unit would support their students in developing an integrated understanding of the disciplinary core ideas, cross cutting concepts, and science or engineering practices targeted in the unit lessons. After teachers made their final customizations, they were invited to share their customized units with other teachers who were interested in implementing the unit. During this unit exchange activity, teachers were able to get feedback from other workshop participants about their customization decisions.

In the final group workshop activity, the researchers demonstrated and helped teachers use the unit authoring tools currently built-in into the WISE platform. During these small workshop sessions, teachers chose 1-2 customizations to reify in the digital version of the unit on the WISE platform.

To conclude the workshop, teachers were asked to provide their reflections on their workshop experiences (see Table 1). Of the 21 teacher participants, 15 (71%) provided responses to the reflection questions. As for whether the KI framework was a productive lens through which to evaluate and customize a unit, 80% of teachers (12/15) answered in the affirmative. In response to the question, “Do you think you could use this process again to customize another unit you'd like to run?”, 87% of teachers (13/15) answered in the affirmative. Additionally, many teachers (87%, 13/15) shared that the lesson series format of the WISE units provide great affordances for customization using their own activities.

## Reflection on and implementation of a customized unit: The cases of Mr. Vega and Mr. Harrison

To continue exchanging expertise after the workshop and to extend the collaborative customization efforts, the first author provided in-class support during implementation of the customized unit. The following are the results of the implementation of the customized Photosynthesis WISE unit (<https://wise.berkeley.edu/project/24548>) in the fall term after the summer workshop by two teachers, Mr Vega and Mr. Harrison. Mr. Harrison worked in the small group that customized the Photosynthesis unit during the summer workshop, however Mr. Vega worked in the small group that customized the Global Climate Change unit. Over the course of implementation of the customized Photosynthesis unit, the first author noticed how both teachers interleaved non-WISE activities, most of which were designed to provide students with additional ideas to supplement their limited prior knowledge about certain topics, like chemical reactions. This observation supports the analysis of the unit customizations made during the workshop where 33% of the total customizations aligned with the Adding Ideas step of the KI process (see Figure 3). In both cases, the teachers used a molecular modeling kit to provide their students with ideas about atoms, molecules, and the nature of chemical reactions related to photosynthesis. Additionally, Mr. Harrison incorporated a multi-modal activity on the conservation of matter during the photosynthesis reaction as a transition from Lesson 1 (Plant Growth Needs) to Lesson 2 (Photosynthesis and Cellular Respiration Reactions). During the post-implementation interview he commented that he used the WISE concept mapping activities to support his students in distinguishing their ideas about energy and matter cycling, the targeted cross-cutting concept in the unit. Mr. Harrison's use of WISE activities over his own activities to help students distinguish their ideas correlates with the result from the customization activity, namely the low levels (10%) of customization activities aligning with the distinguishing step (see Figure 3). This finding also parallels Mr. Vega's reflection comment that he plans to "[s]trengthen the role WISE plays during the Distinguishing...phase".

### Mr. Harrison: A deeper look

During the workshop, Mr. Harrison expressed reticence regarding the customization partnership. His past experiences with the units led him to feel like they were "*adapted to fit the need of each researcher*", additionally he could not see how the digital version of the WISE unit could be restructured into related but independent lessons aligned to specific NGSS performance expectations. However, when asked how he thought the customized, digital Photosynthesis unit aligned with the NGSS he remarked, "*It's definitely aligned, it fits right in with our [other curriculum materials] because we go from Photosynthesis right into ecosystems*" (which is Lesson 4 of the customized Photosynthesis unit). This comment highlights the way that Mr. Harrison interleaved the WISE Photosynthesis unit with his existing curriculum to better support his students in meeting the targeted NGSS performance expectations. Specifically, Mr. Harrison commented that his existing curriculum, although nominally aligned with the NGSS, did not actually provide his students with sufficient opportunities to develop a robust and coherent understanding of the targeted ideas, concepts, and practices. He, therefore, used the customized WISE unit to provide his students with these opportunities. Mr. Harrison further discussed the difficulty and discomfort he experienced when implementing curriculum that he did not develop and therefore greatly appreciated the ability to customize the WISE unit with his own activities to make the unit "his own". When the first author asked Mr. Harrison how the researchers could further collaborate to address his curriculum customization challenges, he offered the idea of incorporating specific WISE lessons and activities into the Google Classroom that he used to plan and organize his instruction. He stated that doing so would solve the notoriously difficult problem of trying to re-synchronize students who progress at different speeds through the WISE activities. (We elaborate upon this idea in the Conclusion section.)

### Mr. Vega: A deeper look

Mr. Vega also engaged in extensive customization during implementation to assume greater "authorship" of the unit. During the implementation, he commented to the first author, "*I was motivated to Edit Content to the unit...Take a look*", referring to his use of the unit authoring tools on the WISE platform, a feature historically used only by researchers. In these edits, Mr. Vega customized the prompts for the concept maps to align with the topics and terminology used in his non-WISE activities. In this way, he created greater continuity between WISE and his other curriculum materials. Mr. Vega's customization of the Photosynthesis unit during implementation is of particular note since during the workshop customization activity he was not part of the Photosynthesis small group. Therefore, Mr. Vega's customization of the Photosynthesis unit demonstrates that he was able to transfer and apply the knowledge he gained during the workshop to another unit, actions that substantiated his "YES" to the reflection question of whether he could customize another unit. Beyond making edits to provide greater continuity between the WISE unit and his existing curriculum materials, Mr. Vega also

applied his understanding of the KI process to all his instruction. During the post-implementation interview he stated that he used the steps of the KI process to keep track of the stage of learning in which his students were engaged and thus provide them with targeted support.

## **Discussion**

### **Design principles for Knowledge Integration customization**

In this paper, we described professional development activities that were aligned with the KI framework and showed how the framework helped teachers customize their existing curriculum to better support their students in meeting the three dimensional learning goals of the NGSS. The workshop reveals three design principles that we recommend for teacher-researcher professional development: support teachers and researchers to learn from each other; make thinking about customization visible; and support sustainable customization practices. In the sections below, we highlight the value of each design principle in supporting collaborative learning amongst researchers and teachers. We conclude the section with a discussion of how the insights gained during the customization activity inform the design of an online curriculum customization interface.

### **Supporting teachers and researchers to learn from each other**

The workshop facilitated the integration of teachers' and researchers' knowledge and expertise. Throughout the workshop activities, researchers used the KI framework as a mediating tool to share their insights into the learning process, and teachers shared their insights into the instructional constraints and resources of their teaching contexts. Transitioning from whole to small group activities allowed workshop participants to learn from each other and consider new strategies and activities to help students meet the learning goals of the NGSS. Teachers expressed that having opportunities throughout the workshop to learn from researchers and other teachers within and outside their school was invaluable, as they do not regularly have such opportunities. The final customized WISE units reflected the integration of the ideas and expertise of researchers and teachers. Thus, the workshop embedded research-based pedagogical insights into teachers' curriculum customization practices and expanded researchers' understanding of teachers' instructional constraints and resources.

### **Making thinking about customization visible**

To customize the WISE units, we designed workshop activities that made the customization process visible. The diagrammed WISE units made the researchers' thinking visible to teachers by highlighting the units' salient features and the researchers' design rationale. The diagrammed units facilitated productive conversation about the feasibility of unit implementation and theories of learning, specifically the KI framework. Writing their ideas for lesson activities on sticky notes color-coded according to the steps of the KI process provided teachers tangible artifacts with which to organize their current curriculum materials in ways that would support constructivist-grounded pedagogy. Having the content of the sticky notes at the grain-size of lesson activity supported experimentation with lesson structure and sequence. Teachers could place and replace the notes on the lesson notecards. Thus, the customization activity helped participants make their thinking visible by moving their ideas from conceptualization to paper, making their ideas available for evaluation and revision by themselves and others. The activities also helped make learning the KI framework accessible to teachers, thereby providing teachers and researchers with a common pedagogical lens with which to evaluate and customize curriculum.

### **Supporting sustainable customization practices**

After the workshop, teachers and researchers partnered to implement the customized WISE units. The two cases presented in this paper, highlight the power of the professional development activities to promote sustainable curriculum customization practices. In one case, the teacher (Mr. Vega) implemented a customized Photosynthesis WISE unit which he did not work on during the workshop. During implementation he applied the KI framework to effectively interleave his existing curriculum activities into the digital WISE Photosynthesis unit to support his primarily English Language Learner students in developing a coherent understanding of the targeted performance expectations. Similarly, the other teacher (Mr. Harrison) customized the WISE Photosynthesis unit to supplement his existing curriculum. He recognized that his existing curriculum did not support students to distinguish their ideas, a critical step in the KI process. He used WISE activities to do so. Both teachers noted that the structure of the WISE units, namely a series of related yet independent lessons targeted to specific NGSS performance expectations that engage students in each step of the KI process, made customization feasible and effective. Rather than viewing the WISE units as a product of research efforts

that needed to remain unedited, both teachers viewed the units as their own and acted accordingly. The workshop developed teacher agency around curriculum customization thereby allowing the customized WISE unit to be the product of a partnership between education experts, experts of theory and experts of practice.

## Conclusion

The outcomes and associated principles presented in this paper illustrate the promise of the KI professional development activities to support the alignment of existing curriculum with NGSS. Analysis of the activities suggest three design principles that echo the KI tenets: learning from others, making thinking visible, making learning accessible, and promoting autonomous learning (Linn & Eylon, 2011). The principles were developed in the context of STEM curriculum customization; however, they should also be tested in other contexts, such as developing culturally-relevant curricula. The findings also illustrate the value of logging student activities and using student work to support customization. Incorporating learning analytics into the WISE units could potentially allow teachers to use evidence of student learning in real-time to make curriculum customization that support the KI process. This is an application of these findings that we are actively investigating.

## Make customization accessible

The diagrammed version of the WISE units and the workshop activities provided a prototype for an online customization and implementation interface. This interface makes the customization process accessible to teachers. Teachers commented that the diagrammed WISE unit provided the ideal amount and type of information they needed to gain a working knowledge of a unit. They also valued the intuitiveness of customizing units with activity-level sticky notes color-coded for the KI process. Using these experiences, the researchers and WISE technology team designed a customization and implementation interface on the WISE platform. When teachers view a WISE unit from the interface they will see a pop-out window with the same information that was on the notecards. Upon selecting the unit for implementation, the unit will be displayed at the activity level with each activity tagged for the KI step that it targets. Teachers can add specific activities from other WISE units or from their own resources. In the future, we anticipate integrating with Google Classroom so that teachers can seamlessly combine their existing curriculum with WISE lessons or activities. Additional analysis of the final customized unit will help ensure that teachers have a customized curriculum that engages their students in complete KI cycles. The interface will allow teachers and researchers to continually benefit from each other's respective expertise and sustain their curriculum customization partnership.

## Next steps

Based on the successful implementation of the customized WISE Photosynthesis unit, we will study the implementation of all the customized WISE units. These findings will allow us to refine the online customization and implementation interface and the workshop design.

## References

- Gerard, L. F., Varma, K., Corliss, S. B., & Linn, M. C. (2011). Professional Development for Technology-Enhanced Inquiry Science. *Review of Educational Research*, 81(3), 408–448. <https://doi.org/10.3102/0034654311415121>
- Lave, J. (1996). Teaching, as Learning, in Practice. *Mind, Culture, and Activity*, 3(3), 149–164. [https://doi.org/10.1207/s15327884mca0303\\_2](https://doi.org/10.1207/s15327884mca0303_2)
- Linn, M. C., & Eylon, B.-S. (2011). Science learning and instruction: Taking advantage of technology to promote knowledge integration. New York: Routledge.
- NGSS Lead States. 2013. Next generation science standards: For states, by states. (2013).
- Penuel, W. R., Fishman, B., Yamaguchi, R., & Gallagher, L. P. (2007). What makes professional development effective? Strategies that foster curriculum implementation. *American Education Research Journal*, 44, 921–958.
- Penuel, W. R., & Gallagher, L. P. (2009). Preparing teachers to design instruction for deep understanding in middle-school earth science. *Journal of the Learning Sciences*, 18, 461–508.

# The Affordable Touchy Feely Classroom: Textbooks Embedded With Manipulable Vectors and Lesson Plans Augment Imagination, Extend Teaching-Learning Practices

Durgaprasad Karnam, Harshit Agrawal, Priyanka Borar, and Sanjay Chandrasekharan  
karnamdpdurga@gmail.com, harshitagrawal.iitr@gmail.com, priyankaborar@gmail.com,  
sanjay@hbcse.tifr.res.in

Learning Sciences Research Group, Homi Bhabha Centre for Science Education, TIFR, Mumbai-India

**Abstract:** Computational media for mathematics and science education have been around for forty years, but teaching and learning practices are still centered surrounding the textbooks in the developing world, mainly due to their affordability. Teachers thus organise their thinking, workflow and classroom interactions using textbooks. However, text media limit students' ability to imagine the dynamics embedded in mathematics and science formalisms, such as vectors. We outline an affordable design that augments textbooks, to: allow all students to imagine such dynamics; help teachers reorganize their thinking and workflow; make all classrooms highly collaborative. The design consists of a 'touchy-feely vector' system (smartphone-based manipulation of vectors), a 'virtual lesson plan' (manipulable simulations that extend textbook figures), and a 'mixed-media' textbook (QR codes pasted next to figures, linked to manipulable simulations). We discuss how this integrated design changed learning and teaching practices, based on an ethnographic study of experimental and control classrooms.

## Introduction

Interactive media for making concepts and formal structures in mathematics and science easier to understand (Balacheff & Kaput, 1996) have been around for close to forty years (Papert, 1980, 1994). But these media have not changed classroom practices significantly, as teaching and learning processes and teacher training are still centred around textbooks, particularly in the developing world. This is because textbooks are cheap and widely available, and they thus play an artifactual role in teaching practice, helping organise teachers' thinking, workflow and classroom interaction. The textbook is thus the dominant media in education, but text and figures are limited in conveying the dynamic nature of mathematics and science content, particularly formal structures such as vectors, which are critical for developing model-based reasoning, a key skill needed for practising science (Hestenes, 2010; Magnani, Nersessian, & Thagard, 2012; Nersessian, 1999). Teachers struggle to enact or 'act out' in the classroom the dynamic nature of such formal structures, and students have a hard time imagining the dynamic nature based on the teacher's gestures, words and enaction and dropout physically or cognitively from the classroom learning.

## Our design TFV-2

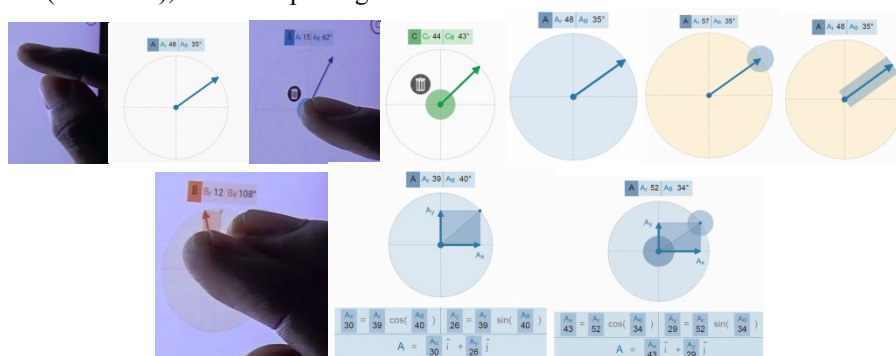
To address this problem for teaching/learning vectors (Barniol & Zavala, 2014; Dorier & Sierpiska, 2002; Hillel, 2000), we developed a multi-pronged design (including cognitive, practice and access factors) to make the dynamic nature of formal structures (geometrical ones in vectors) available to all students, particularly in the developing world.

*Cognitive:* A simulation system was developed, which allows by touching and manipulating them on any smartphone (Touchy Feely Vectors-2: TFV-2: [bit.ly/tfv-2](http://bit.ly/tfv-2)) using *embodied* interactions, for creating (by simple double-consecutive taps), adding (drag one vector over other and double-consecutive taps), and resolving (pinching away gesture) (Fig.1). This makes the formal entities tangible, dynamics visible, and the modes coherent. An earlier version of TFV-2 was tested to be promising in helping students understand and imagine the dynamics better (Karnam, Agrawal, Mishra, & Chandrasekharan, 2016; Karnam, Agrawal, & Chandrasekharan, 2018).

*Practice:* Teacher adoption and usage of digital resources in the classroom has not been smooth (Bingimlas, 2009). In our experience too, teachers were hesitant to use it in their teaching. To address the integration with teachers' current textbook-based practice, particularly their thinking about vectors and the sequencing of the lessons, we decided to design tasks that could be planned for organised classroom activities. As an exercise, a 'virtual lesson plan' (VLP) was created, in collaboration with teachers, where this manipulable system can be used to *extend* their existing teaching lesson plans.



Access: These VLPs were *embedded* in the textbooks through QR codes (Uluyol & Agca, 2012) augmenting the figures and tasks in the textbook, as well as teacher thinking and practices based on them. This makes the introduction of new artefacts into the classrooms smooth and less abrupt as it builds on the already existing artefacts (textbooks), without replacing them.



**Figure 1.** Gesture for creating vector, and for deleting a vector; An active vector, changing magnitude, direction (row-1); Pinching away gesture for resolution and gesture for reversing the resolution of vector (row-2).

We report a study involving classroom teaching using TFV-2 and the changes in teaching and learning practices, based on data from an ethnographic analysis of experimental and control classrooms.

### The classroom study

To test whether this design actually meets the design requirements, we did a classroom study (grade 11), with three experimental classrooms where the integrated system was deployed, and three control classrooms where the standard textbook teaching method was used. We worked with 5 physics teachers (T1-T5), in 3 junior colleges (C1-C3) (pre-university level high schools) in Mumbai, India. All the teachers were trained postgraduates in Physics and had about 20 years of experience teaching at this level. Each school had multiple classrooms (divisions/sections) of grade 11. We observed two grade 11 classrooms (Control group - CG, and Experimental group - EG) in each college (a total of 6 classrooms). In CG, the teachers taught using the conventional method. In EG, the teacher used tablets with the TFV-2 deployed.

**Table 1: Colleges and classrooms, and corresponding teacher codes. NCERT - national, MH - a provincial (Maharashtra) textbook. \* T3 & T5 taught in C2 and C3 due to inter-school transfers (could not be controlled)**

| College (type, Curriculum) | Classroom | Teacher  | No. of students |
|----------------------------|-----------|----------|-----------------|
| C1 (Private, MH)           | EG1       | T1       | 73              |
|                            | CG1       | T2       | 70-75           |
| C2 (Public, MH)            | EG2       | T3       | ~49             |
|                            | CG2       | T3 & T5* | ~ 41            |
| C3 (Public, NCERT)         | EG3       | T4       | 45              |
|                            | CG3       | T5 & T3* | 45              |

### Methodology and data collection

See figure-2. A brief introduction to the research study and the TFV-2 system was given to the respective college managements and the teachers to get their consent for the study. To get some initial familiarity with the system, the teachers were requested to do some exploration of a few tasks (not mandatory) on an earlier (mouse-control) version of the vector system. All the teachers except T5 attended a Lesson Planning Workshop (LPW) for ~3 hours. In the workshop, the teachers based on their already existing lesson plans, together discussed and arrived at a virtual lesson plan with tasks on TFV-2, to be used in their teaching. We video recorded all the classroom sessions (both in CG and EG for 5-8 teaching sessions of 35/40 minutes each in different colleges). In CG classrooms, the teacher was given no inputs by the researchers. In EG classrooms, as per the VLP, a 1-2 page worksheet with pre-determined tasks (developed during the LPW) was prepared, depending on the topics the teacher planned to teach in the class (one worksheet per student). About 25-30 6" Android tablets (SWIPE)



with the necessary files and QR reading applications pre-installed were taken to the classrooms. The classroom was split into groups of 2-3 students (depending on the total strength) and each group was given one tablet. The tasks are available over the web and can be accessed directly. But the colleges didn't have reliable internet connectivity, so the necessary files were copied to each tablet, and they were accessed locally using QR codes generated for the file paths. In the EG session, the teacher led the class, discussing the day's topics. The students in the groups interacted with the tablet-based system, as directed by the teachers and facilitated by the researchers (fig.3). They also completed related worksheets. The researchers' roles were primarily logistics related to the tablets, technical support (for the teacher as well as the students), facilitating activity based on TFV, and recording the classroom proceedings. After about 1-3 weeks, students in all the 6 classrooms were given a questionnaire, which included questions that sought to capture students' conceptual understanding and reasoning approaches in the context of vectors.

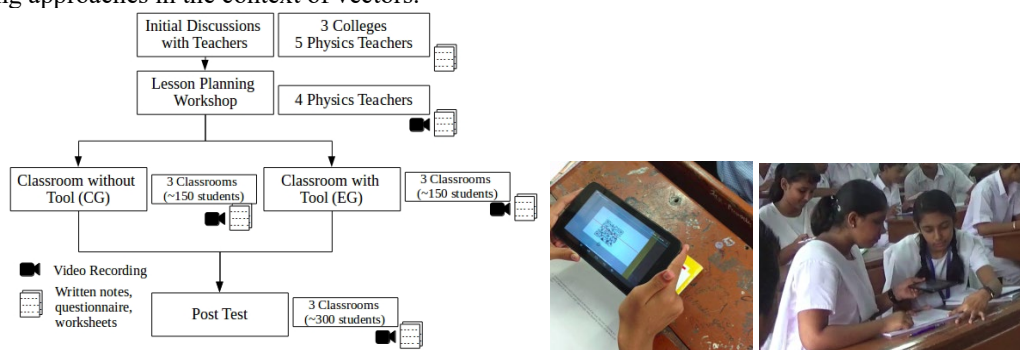


Figure 2. The flow of the study and the sample. Figure 3. Students scanning the QR code in an EG Classroom.

## Analysis and findings

Given the space constraints, we confine to reporting here the analysis of qualitative changes, both in students' learning (particularly imagination) and teachers' teaching practices in the classroom setting. The main data sources for the current analysis were video recordings, which were analysed to understand the differences in: classroom dynamics (CG and EG), interactions between teachers and the students, and interactions with various artefacts (textbook, blackboard, notebooks and worksheets, the tablet). For this, the video recordings were analyzed using two broad lenses – differences in learning, particularly imagination, and differences in teacher/teaching related practices. This is not a rigid and complete framework (like situated frames of instrumental genesis (Artigue, 2002) or socio-mathematical norms (Yackel & Cobb, 1996) etc) yet, but, we expect that iteratively useful holistic frames could emerge to analyse the new mixed media based classroom artefacts perhaps enhancing the existing frames incorporating embodied indicators of imagination like gestures.

## Learning practices

The main differences in the learning process, particularly related to imagination, between the CG and EG classrooms were in terms of the use of gestures, trajectories, collaboration and flow of learning.

### Gestures: Active and meaningful engagement with content

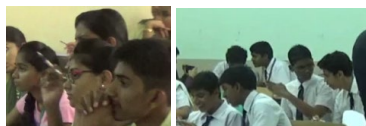


Figure 4. Student in CG using gestures to remotely access the content on the board (left). Student in EG using gestures to explain to the peer in the context of an activity on TFV (right).

As one of our key design objectives was to help students imagine the vector in geometric terms, we examined the gestures students made in the two classrooms, as these indicate geometric thinking (Kim, Roth, & Thom, 2011). In the CG classroom, students were passive spectators most of the time, except for a few episodes where they took note of what was on the blackboard. Activity, when rarely present, was limited to nodding in agreement or tracing pens over the diagrams seen on the board (fig. 4 left). Gestures of this kind suggest a very limited engagement with the geometric content of vectors. The teacher's enaction of the textbook figures, which is the sole source for triggering the imagination of dynamic operations, is not resonated by the students. In the EG classrooms, more active and meaningful engagement with the content was observed. This was reflected in

more sustained gestures, along with discussions between students, which indicate conceptually deeper engagement with the geometric operations (fig. 4 right). These gestures were usually situated in conversation, where one student was actively trying to explain his imagination to another. The gestures indicate the possibility of students resonating with the teacher's writings (equations) or drawings (diagrams) and meaningfully engagement with content.

### Multiple trajectories for learning

In the CG classroom, the learning opportunities were only when the teachers enacted a concept, and most of the activity in the classroom involved taking notes, which followed the teacher's description of the topic and solutions to problems. In the EG classroom, the exploratory and open-ended nature of tasks led to a surprising number of problem-solving trajectories. The diversity of approaches indicate the high imagination and learning potential provided by the TFV-2. To illustrate, in one of the tasks, students had to create two vectors, whose resultant is a given vector with magnitude 60 and a direction of  $40^\circ$ . The following approaches were seen.

- Trial and error (the most common approach, where students manipulate the magnitude and direction of vectors randomly). Some students found some patterns of change as they kept interacting.
- Estimating the rectangular components along x and y-axes and creating them as the two vectors. However, these students struggled when asked to create another set of vectors. Here the students explicitly applied the interconnections between resolution into rectangular components and addition.
- Creating two vectors with magnitude 30 at  $20^\circ$  from the x-axis and then adding to find resultant of magnitude 60 at  $20^\circ$  instead of at  $40^\circ$ . This combination provides the opportunity for students to realise that magnitudes and directions (represented as angles) cannot be algebraically added directly.
- Creating the vector of magnitude 60 at  $40^\circ$  from the x-axis, and arguing that the other vector is a zero vector. This particular group of students tried out this task for a long time using trial and error method, and eventually applied the idea of zero vector quiet intelligently.

### Collaboration

In the CG classroom, there is very limited scope for students to engage in discussions with peers. The only conversations recorded between students involved asking for an extra pen or pencil, or to show some part of his/her notes for clarification, or giggles with playing (disinterested in the content of the lesson).

In the EG classroom, the very nature of the lesson plan involved making groups of 2 or 3 students, who worked together on the tablet. This inherently made the class collaborative. In addition, there were many episodes (similar to fig.5) where students naturally started interacting across groups. The nature of collaboration was grounded in the tasks they were performing on the tablets. Collaboration is not necessarily an indicator of an effective classroom, as there could be a lot of social and cognitive loafing (O'Donnell & O'Kelly, 1994), which could have happened in the EG classroom as well. However, as the students were involved in the collaborative activity using the TFV-2, the teacher could monitor student progress, to start discussions.



Figure 5. Students in EGs collaborating within their designated groups as well as beyond their groups.

### Control on the flow of learning

In CG classrooms, most of the time was spent on the teacher delivering content in the lecturing mode. The teacher thus had full control over the flow of knowledge and learning in the classroom. There was very little scope for students to intervene in the flow and build their own knowledge. Further, the teacher decided when students could take notes, when to listen to him/her, which book was to be taken, who should respond to questions etc. This authority stance is central to the conventional classroom in developing countries, and this is required to some extent for smooth classroom management. However, this power structure often makes students intimidated, and they thus seldom interact with the teacher. Students tend to overcome this fear using chorus responses. The teacher wielded further control by appreciating (saying words like very good, interesting) or dismissing (ignoring or not attending to) student responses.

EG students actively participated in the enaction process along with the teacher, and this transferred agency to the students, who had more power in directing the flow of learning than in a normal classroom. This was reflected in the free interactions among students, the change in the nature of teacher-student interactions,

and also students' emotional connections with the tasks. The students were active participants with the teacher, and their interaction with the content was active and meaningful (see fig.6) as can be seen in the nature of gestures used. Based on these interactions, along with questions and discussions with the teacher, they controlled the flow of learning, and hence constructed their own knowledge. The teacher was no longer the sole controller of the flow of knowledge. Students generated knowledge during their interactions with their peers mediated by artefacts like TFV-2, and knowledge flowed horizontally through collaborative work. Moreover, as suggested by the teachers, if tasks could be designed for students to do at home (thus flipping the classroom), the flow of learning could become even more student-centred.

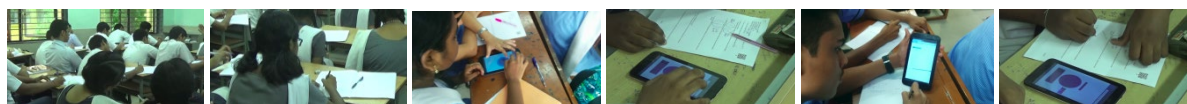


Figure 6. Students in a CG taking notes in their notebooks (1,2) Students in EG interacting with the tablet (3,4) and taking note of it in the worksheet (5,6).

### Other observations

The CG classrooms were very calm, and there were never any emotionally charged moments. In contrast, in the EG classroom had a charged feel, and there were numerous moments of excitement and disappointment. There were multiple instances where EG students asked for similar systems for other topics in the curriculum. Students were very excited to see the tablets on the first day. By day 3 and 4, they started engaging with the TFV-2 in a more serious way and were not just excited due to the presence of the tablet.

### Teacher/Teaching practices

The introduction of the TFV-2 into the classrooms was not fully and easily embraced by the teachers, though there were changes that indicated possible integration eventually. Here we discuss some of the central differences (and lack of changes) captured in the video data. We start with a discussion of practice elements resistant to change.

#### Resistant teacher practice

Three out of 5 EG teachers (T1,T2,T5) used the same diagram for teaching parallelogram law of vector addition, where the angle is less acute, even though the vectors used in triangle law, shown simultaneously on the board, were quite different (See fig.7). This drawing closely follows the textbook one and indicates the deep and subtle influence of the textbooks, where the teachers are conditioned by the textbook representation to such an extent that the diagrams seep into their practice without explicit awareness. While the TFV-2 has changed the way teachers present the content (as discussed in later subsections), there were such subtle aspects that were still driven by the textbook representations. We present this result first to indicate how deeply the textbook is integrated into teaching practice. This conditioning by textbooks extended to the way content was presented. In the EG classroom, where teachers had the possibility of widening the scope of discussions (like emphasizing and discussing the patterns of changes in the algebraic expressions through geometric manipulations, or the nature of the  $x$  and  $y$  components being interconnected by the circle). Such topics were not initiated by the teacher, and the scope of the class was still within the limits of the topics presented in the textbook. This reaffirms the conditioning role played by textbooks in setting classroom practices.

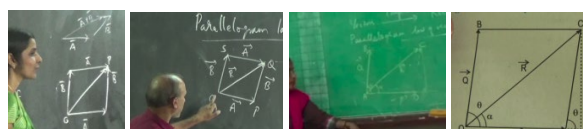


Figure 7. Teachers making parallelograms very similar to the one given in their textbook (right) with almost  $90^\circ$  angle while teaching parallelogram law of vector addition.

In the CG classroom, the teaching narrative usually took the form of a lecture, where the teacher introduced the definition or the statement of a law, explained it using some examples, and then drew diagrams. The teacher used some gestures to enact the diagram. The students took written notes based on these practice elements. When some extra time was available before the end of the session, the teachers solved some numerical problems. The flow of the lesson was very linear and monotonic. In the EG classroom, the narrative was similar in the beginning (introduction to the definitions or statements, making diagrams). It then changed to describe the tasks the students needed to do using the system. But once students started interacting with TFV-2, this was

altered. Interestingly, there was a tendency for both CG and EG teachers to follow a set question and answer template for discussion (especially when stating the definitions or the laws of additions), as this helps students in standard assessment. In later discussions, T1, who has taught in an EG class, noted that despite the teachers being given the flexibility to use new modes of teaching, there is a lot of emphasis on examination results and time limitations. This emphasis comes from the way the education system in India is structured around exams, which restricts the possible narratives in the classroom.

### Practice elements that changed

A central change was the nature of teacher talk in EG classrooms, from mere description of static diagrams (like one arrow and another arrow forming two sides of a triangle, and the third side being the resultant, in triangle law of vector addition), to a more dynamic enactment, where teachers described the addition process. For instance, the teacher used the gestures in the TFV-2 to explain the process of adding two vectors saying “*now you know how to take vectors, just bring this vector (2nd vector) and attach here (head of 1st vector), then you will have to press your thumb (at the tail of 1st vector) and other finger here (head of 2nd vector), you will get the resultant*”. Teachers in both CG and EG tended to use many gestures and bodily actions when explaining content using diagrams. However, there were differences in the gestures used by CG and EG teachers. In CG, teachers tended to use iconic gestures (like drawing an arrow in the air or using an opened palm to represent a vector) which are linked to diagrams on the board. Some were metaphorical (Roth, 2001) gestures (such as hand movements to show the splitting of vectors during resolution). In the EG classroom, teachers used similar gestures as a CG teacher during lecturing. However, they also incorporated gestures used in the TFV-2. For example, T4 and T3 used gestures influenced by the system when explaining triangle law of vector addition (see quote above). These gestures signify a dynamisation of the vector operations, based on interactions in TFV-2

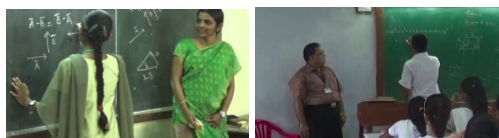


**Figure 8.** (Left) Teacher in CG using iconic gestures to show a vector, which is still a diagram. (Right) Teacher in EG using a picking action, reflecting that the vector is a touchable entity.

In the CG classroom, the teacher stayed near the blackboard (as seen in many figures). As the TFV-2 allowed all EG students to enact vector operations, the teacher did not need to demonstrate the enactment in the class, and she thus had a lot of time available to move around in the class and engage in discussions with the students. The role of the TFV-2 in generating this behaviour is clear from the contrasting behaviour of T3 (only teacher present in both EG and CG) in the CG and EG classrooms. The discussions involved clarifying conceptual doubts and questions. This allowed the teacher to also get real-time feedback about students’ understanding of vector concepts, and initiate discussion with individual groups or the entire classroom.



**Figure 9.** (Left and Center) Teachers engaging in discussion with student groups in EG. (Right) Teacher showing a group’s work on the tablet to the entire classroom (a potential space for initiating a discussion).



**Figure 10.** Teachers in CG calling students to the board to answer or solve some questions.

Usually, teachers ask questions to check for understanding (Rosenshine, 1983), or when students are not attentive. In the CG classroom, the monologue of the teacher was often interspersed by questions, where the teacher explained a concept or a problem and asked one of the students to stand or come to the board and solve or answer a similar problem (Fig.10). There were some such episodes in the CG classroom, where the teacher called a student to perform subtraction of two vectors, or derivation of expression of parallelogram law by simple substitution, or to solve a numerical problem similar to the one the teacher had just solved. Irrespective

of the purpose of asking the question, they were close-ended questions with a unique answer.

In contrast, in the EG classroom, the teacher spoke to the entire classroom only in the beginning. The rest of the time was spent answering questions from students. The nature of the questions was open-ended (with no single answer) and exploratory, following the exploratory nature of TFV tasks. Besides the above changes, a few other episodes indicate significant changes in teacher thinking and practice, induced by the TFV-2.

As teachers started discussing possible lesson plans during LPW, they immediately suggested that we could design tasks for students to do at home (as they get more time to engage with the TFV-2) and then discuss their findings and responses in the classroom next day, which was akin to flip-classrooms. After one of the lessons, T3 said - "besides helping the students in visualising and understanding the concepts, it (TFV-2) has clarified and made imagination easier for topics for me too". This shows how systems like TFV-2 could also help teachers understand complex concepts better, and also help develop pedagogical content knowledge (Shulman, 1986), particularly given the acute shortage of good teacher trainers and programs in the developing world. The teachers highly appreciated the TFV-2 when it was first shown to them. As discussed, some of the teachers were not comfortable using it in their teaching. However, after a few classes based on teaching using the system, the teachers became more comfortable and confident. Teachers have repeatedly requested extensions to the TFV-2, particularly to include 3D vectors, and also products of vectors, as well as waves. This is because it is difficult to draw and explain these using the blackboard. A physics teacher who was not part of the study sat through an entire classroom session, to see how the system was being used by students. Also, T5, who could not be part of the LPW, and who taught only in the CG classroom, repeatedly requested deployment of the system in the CG classroom. These indicate a shift to a motivation to use technology from an initial hesitation in teachers.

## Discussion

The TFV-2 provides a new model to address the problem of augmenting science and mathematics teaching and learning practices, particularly in the developing world. The focus on smoothly extending textbooks and related practices, rather than replacing them, makes our approach different from other similar approaches to integrating multiple external representations (Bodemer, Ploetzner, Feuerlein, & Spada, 2004), such as GeoGebra. Also, the high scaling possibilities provided by the teacher-driven QR-based augmentation of textbooks offer a new approach to address key resource limitations in developing world classrooms, in terms of both technology access and training costs. The QR-based approach also provides a way for educational research institutions to curate good learning resources (such as videos and problem banks), and include these periodically as annotations to current textbooks, without waiting for the lengthy and bureaucratic process of textbook revision.

We could see a transformation of the classroom space for the students and situations and practices for their learning emerged in them smoothly. Among teachers, we see they could adapt the technological intervention smoothly, but the changes in their practices were far subtler, as we see indicators of them still anchored to textbooks. It is here, a key insight emerges about the central role textbooks play in directing teacher practice, and how building on the textbook could lead to rapid changes in teaching practices, particularly in terms of the way teachers direct/control the flow of knowledge in the classroom. The augmentation of textbooks could also help in swiftly upgrading teacher thinking and pedagogical knowledge of science and mathematics concepts, which is a critical issue in the developing world. Such an upgrade of pedagogical content knowledge (Van Driel, Verloop, & de Vos, 1998) could be better facilitated by QR augmentations that support flipping the classroom, as this feature would raise the number of potential questions from students, and in turn, motivate teachers to come prepared to answer these questions.

The narrow scope of our vector application, which seeks to just seed imagination of geometry, is different from most other interactive systems for learning vectors (such as PHeT), which are scoped more widely. This scoping makes such systems unsuitable while trying to extend teachers' current textbook-based practices, which tend to be very modular. The touchable and manipulable vector also potentially creates a new and embodied way of understanding vectors and related concepts for learners, as has been argued in the case of touchable numbers (Sinclair & de Freitas, 2014). The embodied interaction with vectors also makes our approach different from augmented reality approaches, which focus more on the visual component.

More broadly, our mixed media design stance -- based on an integration of cognitive, practice, and access factors -- treats textbooks as a very useful resource for advancing technology-based solutions for education, particularly in the developing world. This approach seeks to create a bridge from current practice to future practice, where digital platforms such as NetLogo could support the learning of cognitive skills not supported by textbooks, such as computational thinking (Dickes, Sengupta, Farris, & Basu, 2016).

However, we take a more careful design approach towards this shift, analysing the positive role textbooks could play, through their ability to promote working memory based operations such as mental rotation and modelling, particularly by *limiting* dynamic interaction. We believe this reflective stance, and the specific



design instantiation of this stance outlined here provides a way to reevaluate key assumptions underlying educational technologies for learning science and mathematics.

## References

- Artigue, M. (2002). Learning Mathematics in a CAS Environment: The Genesis of a Reflection about Instrumentation and the Dialectics between Technical and Conceptual Work. *International Journal of Computers for Mathematical Learning*, 7(3), 245.
- Balacheff, N., & Kaput, J. J. (1996). Computer-Based Learning Environments in Mathematics. In *International Handbook of Mathematics Education* (pp. 469–501). Springer Netherlands.
- Barniol, P., & Zavala, G. (2014). Test of understanding of vectors: A reliable multiple-choice vector concept test. *Physical Review Special Topics - Physics Education Research*, 10(1), 010121.
- Bingimlas, K. A. (2009). Barriers to the successful integration of ICT in teaching and learning environments: A review of the literature. *Eurasia Journal of Mathematics, Science & Technology Education*, 5(3).
- Bodemer, D., Ploetzner, R., Feuerlein, I., & Spada, H. (2004). The active integration of information during learning with dynamic and interactive visualisations. *Learning and Instruction*, 14(3), 325–341.
- Dickes, A. C., Sengupta, P., Farris, A. V., & Basu, S. (2016). Development of Mechanistic Reasoning and Multilevel Explanations of Ecology in Third Grade Using Agent-Based Models. *Science Education*, 100(4), 734–776.
- Dorier, J.L., & Sierpinska, A. (2002). Research into the Teaching and Learning of Linear Algebra. In D. Holton, M. Artigue, U. Kirchgräber, J. Hillel, M. Niss, & A. Schoenfeld (Eds.), *The Teaching and Learning of Mathematics at University Level: An ICMI Study* (pp. 255–273). Dordrecht: Springer Netherlands.
- Hestenes, D. (2010). Modeling Theory for Math and Science Education. In R. Lesh, P. L. Galbraith, C. R. Haines, & A. Hurford (Eds.), *Modeling Students' Mathematical Modeling Competencies* (pp. 13–41). Springer US.
- Hillel, J. (2000). Modes of Description and the Problem of Representation in Linear Algebra. In J.-L. Dorier (Ed.), *On the Teaching of Linear Algebra* (pp. 191–207). Dordrecht: Springer Netherlands.
- Karnam, D., Agrawal, H., Mishra, D., & Chandrasekharan, S. (2016). Interactive vectors for model-based reasoning. In Chen, W. et al. (Eds.), *The Workshop Proceedings of the 24th International Conference on Computers in Education, India: Asia-Pacific Society for Computers in Education*, 401–406.
- Karnam, D., Agrawal, H., & Chandrasekharan, S. (2018). “Touchy Feely Vectors” changes students’ understanding and modes of reasoning. In Yang, J. C. et al. (Eds.), *The 26th International Conference on Computers in Education. Philippines: Asia-Pacific Society for Computers in Education*.
- Kim, M., Roth, W. M., & Thom, J. (2011). Children’s gestures and the embodied knowledge of geometry. *Journal of Science and Mathematics Education*, 9(1).
- Magnani, L., Nersessian, N., & Thagard, P. (2012). *Model-based reasoning in scientific discovery*. Springer Science & Business Media.
- Nersessian, N. J. (1999). Model-Based Reasoning in Conceptual Change. In *Model-Based Reasoning in Scientific Discovery* (pp. 5–22). Springer US.
- O’Donnell, A. M., & O’Kelly, J. (1994). Learning from peers: Beyond the rhetoric of positive results. *Educational Psychology Review*, 6(4), 321–349.
- Papert, S. (1980). *Mindstorms: Children, Computers, and Powerful Ideas*. NY, USA: Basic Books, Inc.
- Papert, S. (1994). *The children’s machine: Rethinking school in the age of the computer*. Basic books.
- Rosenshine, B. (1983). Teaching functions in instructional programs. *The Elementary School Journal*, 83(4), 335–351.
- Roth, W. M. (2001). Gestures: Their Role in Teaching and Learning. *Review of Educational Research*, 71(3), 365–392.
- Shulman, L. S. (1986). Those Who Understand: Knowledge Growth in Teaching. *Educational Researcher*, 15(2), 4–14.
- Sinclair, N., & de Freitas, E. (2014). The haptic nature of gesture: Rethinking gesture with new multitouch digital technologies. *Gesture*, 14(3), 351–374.
- Uluoyol, C., & Agca, R. K. (2012). Integrating mobile multimedia into textbooks: 2D barcodes. *Computers & Education*, 59(4), 1192–1198.
- Van Driel, J. H., Verloop, N., & de Vos, W. (1998). Developing science teachers’ pedagogical content knowledge. *Journal of Research in Science Teaching*, 35(6), 673–695.
- Yackel, E., & Cobb, P. (1996). Sociomathematical Norms, Argumentation, and Autonomy in Mathematics. *Journal for Research in Mathematics Education*, 27(4), 458–477.
- Young, E. &. (2016). *Future of Digital Content Consumption in India*. EY.

# Evaluating an Adaptive Equity-Oriented Pedagogy on Student Collaboration Outcomes Through Randomized Controlled Trials

Andrew Estrada Phuong, University of California Berkeley, [andrew.e.phuong@gmail.com](mailto:andrew.e.phuong@gmail.com)  
Judy Nguyen, Stanford University, [judynguyen@stanford.edu](mailto:judynguyen@stanford.edu)

**Abstract:** This study evaluates an adaptive equity-oriented pedagogy (AEP) through a randomized controlled trial. AEP employs evidence-based teaching practices (e.g., formative assessment, universal design for learning) to address college students' diverse learning needs (i.e., their strengths, interests, and areas for growth). AEP provides collaborative computer-supported project-based opportunities that enable students to extend the application of course concepts to embedded contexts that engage them as learners. To compare AEP and active-learning control conditions, this study utilizes identical measures: validated surveys, interviews, observation notes, anonymous course-feedback forms, and formative and summative assessments. Multivariate regression analyses suggest that students learning through AEP outperformed control conditions by a full letter grade and scored on average 14.20 percentage-points higher on final assessments, when controlling for disability status, gender, and pre-test achievement. AEP has a large standardized effect size on the final ( $d=2.40$ ) and provides a framework to improve student success through embedded and extended learning opportunities.

## Adaptive equity-oriented pedagogy (AEP), research questions, and rationale

This study evaluates an adaptive equity-oriented pedagogy (AEP) that seeks to improve diverse students' success through computer-based formative assessment and computer-supported collaborative project-based learning. Applying AEP, instructors administer weekly computer-based student assessment and surveys to diagnose students' learning needs (i.e., their strengths, interests, and areas for growth). Using student data and analytics, instructors iteratively adjust how they provide feedback, model key skills needed for summative assessments, and implement deliberate practice activities. Students pursue their interests by applying course concepts to a novel context through collaborative project-based learning.

AEP seeks to engage learners in an embedded space by making the content and context relevant so that students can extend their learning into the real-world as a scientist. These embedded learning opportunities are beneficial since instructors address skills that are relevant to students' lived experiences and contexts outside of the classroom. To achieve this goal, students' learning is embedded in the real world since they can apply concepts to a range of contexts, which include researching longitudinal data on students' mental health, business management practices, and cancer treatments. Through project-based learning, students apply design-based research to collaboratively examine and address shared problems. Students work in teams to develop competing hypotheses based on existing literature, their experiences, and the concepts presented in class. They then gather data to test these hypotheses through statistical software as they collaborate online to collect data, build databases, generate code, produce visualizations, analyze data, and interpret findings. Students work together to extend course concepts in embedded contexts to make sense of real-world data. Collaboration in these contexts is critical, since students offer diverse perspectives on ways to refine research design and analyses. AEP facilitates this process because it identifies barriers to student collaboration. For example, instructors use formative assessment and surveys to diagnose students' learning needs with respect to collaboration.

Through these strategies, AEP enables instructors to overcome barriers that otherwise inhibit collaborative learning in university classrooms. One such barrier is the conflict that arises in group projects when members do not complete tasks on time. Reasons for incomplete tasks include lack of proper planning, miscommunication, students not knowing how to apply concepts to large tasks, and students working on tasks that do not engage their interests and strengths. The AEP model helps students develop a structure for collaborative work. For instance, students form teams based on common interests and are asked to collaboratively create a task chart via Google sheets or other mediums to assign roles during group projects. To create a system of accountability, students use this chart to designate individuals who work collaboratively on tasks that have deadlines. To foster trust and motivation, group members assign tasks and roles that leverage their members' strengths and interests. Groups also co-construct norms to navigate conflict and ensure accountability. In addition, instructors continuously model collaborative strategies and leverage student data to provide additional support that groups need. Although these approaches may be common in K-12, they are fairly uncommon in U.S. higher education (Phuong et al., 2017). By equipping students with the skills to collaboratively address problems using technology, AEP aims to increase students' sense of belonging in STEM, since they are developing skills that

society values and needs to thrive in a digital age. These experiences provide opportunities for students to practice academic discourse and to connect it to their interests and sociocultural identity (Gee, 1996).

To formally study whether AEP can have strong impacts in a university statistics course, we conducted a mixed-methods randomized controlled trial (RCT). We examine underrepresented minority (URM) student groups that are often overlooked in research on higher education pedagogy. Our research questions below examine an AEP treatment course's impact relative to an active-learning control course and on subgroups (i.e., disability status and gender), since these URM groups have experienced low achievement, persistence, and retention in STEM (McGregor et al., 2016; Stout & Wright, 2016). This study controls for students' pre-test achievement on the final assessment in treatment and control conditions, since prior achievement significantly influences subsequent achievement (Hattie, 2009). Based on these interests, we explore the following research questions:

1. Is there a mean difference on final collaborative assessments between students in the AEP treatment condition and students in the control condition, after controlling for disability status (i.e., students with and without disabilities), pre-test achievement, and gender (i.e., gender non-conforming, female, male)?
2. Is the impact of AEP on student assessment scores consistent across disability and gender groups?

Our research questions evaluating AEP are significant since universities experience difficulties recruiting and retaining STEM students, especially URM (U.S. Department of Education, 2017). Research shows that 90% of students leaving STEM cited poor teaching as a primary concern (Seymour & Hewitt, 1997). Specifically, university statistics courses, a foundational requirement for STEM fields, are typically computation-heavy and lecture-based (Allen et al., 2012). With limited opportunities for real-world application, students can struggle seeing the relevance of learning statistics (Allen et al., 2012). Furthermore, URM students (e.g., students with disabilities, gender non-conforming students) with limited mathematical preparation often experience math-phobia and poor academic achievement (Tishkovskaya & Lancaster, 2012) which can affect their confidence, scientific identity, persistence, and retention (Peters, 2014; Stout & Wright, 2016). Since teaching can directly impact student success (Condon et al., 2016), studying pedagogy to improve students' academic achievement, especially for URM, in STEM is critical.

## Potential significance

Increasing recruitment and retention in STEM, especially for URM, is important because this population of students can offer diverse perspectives on problems and solutions. Increasing diversity can enable future leaders to be more responsive to their constituents and environment. Additionally, many industries in today's workforce require competency with computer-based technologies and collaboration in team settings. However, many higher education STEM courses are not adequately equipping students, especially URM, with technology-based competencies and interpersonal, collaborative skills (Gasiewski et al., 2012). This is important since students in a scientific community need to learn to engage with others, recognizing community guidelines and norms. Many of these norms are often not explicit to students, especially for URM and first-generation students who did not have access to academic-oriented scientific communities. AEP seeks to make these guidelines and norms explicit by modeling these norms and providing students with opportunities to collaboratively practice and build on these norms with a community of learners. To draw students into a scientific community, the AEP model assumes that like scientists, students often have an interest and purpose to pursue a research topic. Based on this premise, the AEP model seeks to understand learners' sociocultural experiences and socio-academic interests. Through surveys and formative assessments, AEP enables instructors to understand their students' sociocultural environment and context—their lived experiences and interests that inspire them to pursue STEM. For example, some students are drawn to medicine because their parents suffered from medical conditions and could not address their illnesses due to financial constraints. Hence, AEP empowers instructors to identify what brings students to academic STEM spaces by providing frameworks to understand learners and their sociocultural histories.

As students bring themselves into their projects, we study how students offer diverse strengths and perspectives to practice “thinking scientifically” as they design research questions, make and test predictions, solve problems, and critique their own and others' reasonings (Deslauriers et al., 2011). Students also collaborate synchronously and asynchronously on their computers through Google docs where they work together to share visualizations and to analyze and interpret data. By cultivating environments that mirror a professional scientific community, students have opportunities to collaboratively practice and engage with scientific discourse, where they can embody communicating, behaving, thinking critically, and solving problems like a scientist (Gee, 1996).

Our research represents a significant contribution. First, it is useful for practitioners since it provides a research-based pedagogical framework and tools that instructors can use to increase equity in student outcomes, especially for URM. Second, this study's methodological approach advances higher education pedagogy scholarship practices; this is the first study to our knowledge that uses a randomized controlled trial (RCT) with



the same university instructors where the control condition employs active learning in a university STEM course. In many higher education STEM studies, the control condition differs in multiple respects than the treatment making the source of different outcomes difficult to isolate. For instance, the control is often lecture-based and employs instructors with different characteristics than those in the treatment. In this study, the treatment and control conditions are more precisely contrasted allowing us to focus more precisely on AEP effects of interest.

## Theoretical approaches

### Conceptual framework of adaptive equity-oriented pedagogy (AEP)

In this section, we describe AEP's key elements and the learning theories that underpin them. Drawing on McCallum's (2013) Assessment-Instruction (A-I) conceptual framework, AEP focuses on collecting ongoing diagnostic assessment data to guide classroom instruction, decisions, and lesson planning. These assessments gather information on students' backgrounds, interests, and experiences to understand the cognitive and noncognitive factors that impact student engagement in light of their sociocultural history. Applying AEP, instructors use weekly data on students' learning needs to continuously adjust how they

1. foreground how course concepts are relevant to students' goals and shared contexts
2. provide brief warm-up active learning exercises where students can engage in productive struggle with concepts and make meaning individually or collaboratively
3. model software skills, expert thinking on concepts, and strategies that students need to excel on summative assessments; during this time, instructors build on and respond to students' critical thinking strategies from previous assessments and/or productive struggle experiences
4. provide written steps and strategies (e.g., task analysis) of how to approach problems that align with rubrics and the rigor of summative assessments
5. include class time for students to practice these skills and strategies to collaboratively analyze and interpret data on computers; during these deliberate practice activities, students articulate to each other which concept-driven step-by-step strategies are useful for their team final project and why
6. provide low-stakes feedback during class to address misconceptions and close gaps in understanding
7. offer in-class and project-based opportunities for students to incorporate feedback and reinforce learning

AEP seeks to support URM students' success by building on McCallum's (2013) A-I model and Vygotsky's (1978) sociocultural theory of learning, in which learning is situated as a cultural and social process. To avoid the lack of social, collaborative interaction associated with more traditional lecture-based pedagogies, AEP draws on formative assessment and universal design for learning (UDL) to address a wide range of learning needs. Throughout this process, AEP provides an adaptive active-learning framework that helps instructors use formative assessment to iteratively adjust their UDL and collaborative learning practices.

### Formative assessment

AEP uses formative assessment (Black & Wiliam, 1998) by engaging Vygotsky's (1978) sociocultural theory and the zone of proximal development (ZPD), where an instructor assesses a student's actual developmental level (i.e., a student's level without assistance) to help them reach their potential developmental level (i.e., a student's level with assistance). Using formative assessment data, the instructor modifies evidence-based practices (e.g., modeling key skills, collaborative deliberate practice opportunities, feedback) to help students reach that potential. Instructors assess and provide feedback to all students via formative assessment instead of assuming that students enter the class with the appropriate resources and skills to meet course goals. Furthermore, formative assessment aligned with the final exam's rigor clarifies expectations for students unfamiliar with the dominant cultural capital and academic discourse assumed in many college classrooms (Bourdieu, 1973; Phuong et al., 2017). Formative assessment promotes equity in student outcomes since struggling students are more likely to persist in a field when they develop stronger conceptual foundations, reflect on their growth, have a collaborative community, and are confident in their work (Ambrose et al., 2010; Phuong et al., 2017).

AEP's in-class formative assessment, modeling, collaborative practice, and feedback loops promote equity in student outcomes since they support novice learners and underrepresented students who may struggle with challenging concepts. According to Schwartz et al. (2015), novices "find it harder to engage in deliberate practice on their own, because novices often do not have the experience to know what skills they should be working on or how to go about it" (p. 295). Incorrect types of practice can reinforce misconceptions. AEP's in-class formative assessment, modeling, practice, and feedback loops allow instructors to review challenging concepts. AEP also leverages collaborative learning via peer-to-peer support since peers who more recently

learned concepts often empathize with barriers to understanding concepts. Often, these peers can break down course material in digestible ways for struggling and advanced students. This form of reciprocal teaching enables students to work with their peers and draw on each other's expertise to move farther together in their ZPD.

In AEP, modeling key skills and strategies provide opportunities for instructors to offer various worked examples that show students "what to do and why" when they approach problems (Schwartz et al., 2015, p. 295). Instructors applying AEP provide step-by-step explanations for solving problems in writing so that students can understand different ways to read prompts, identify underlying concepts in the problem, explain why they are doing each step, articulate how each step connects to key concepts, and understand when these steps can and cannot be generalized to other situations. These step-by-step strategies further reduce cognitive load because the steps help chunk students' mental processes and deliberate practice opportunities (Sweller, 1994). This approach can make learning new concepts less daunting for novice learners. These strategies and collaborative practice opportunities can help improve students' success and self-efficacy with mastering challenging concepts, especially for URM students who had less access to college-level coursework. AEP seeks to improve students' conceptual understandings and self-efficacy so they can actively contribute to their project teams and their peer's learning during computer-supported, collaborative activities.

These formative assessment, collaborative practice, and feedback loops can also clarify expectations for success and foster students' metacognition and self-regulation processes (Nicol & Macfarlane-Dick, 2007). For example, students reflect on and monitor whether they understand concepts when they take formative assessments and practice scientific thinking. This process can help students identify their strengths, interests, and areas for growth, which can help them set academic goals and develop strategies to advocate for their learning. For URM students who feel underprepared for college STEM courses, providing weekly ungraded assessments and supportive collaborative activities creates a psychologically safe space. Creating a psychologically safe space is crucial because many URM students have experienced high incidences of microaggressions and low levels of STEM achievement, sense of belonging, persistence, and retention rates (Stout & Wright, 2016). Offering greater opportunities for URM students is necessary because many URM families often do not have equitable access to extracurriculars, resources, and rigorous curricula compared to more privileged families (Martin et al., 2016).

### Universal design for learning

To support diverse learners, AEP also incorporates UDL to optimize learning by engaging affective, recognition, and strategic brain networks (Rose et al., 2002). UDL enables students to learn, demonstrate, and reinforce knowledge in various ways by providing multiple means of engagement, representation, and action and expression (Rose et al., 2002). For example, AEP instructors synthesize lectures, simulations, technology, computer-based activities, and project-based learning. Building on Lee's (2005) culturally-responsive-teaching framework, instructors apply UDL strategies to draw on students' funds of knowledge—their background, interests, aspirations (Moll et al., 1992)—to increase their sense of belonging and engagement with developing academic skills. To increase student engagement with academic content, instructors: 1) use student survey data to explicitly articulate how students' interests and aspirations are relevant to course material during interactive lectures and active learning activities, 2) visualize challenging concepts with written annotations, and 3) ask students to apply course content to their own contexts. Consequently, students can see how their backgrounds, interests, and identities are relevant to and are strengths in STEM. In addition, students can exercise agency as they apply concepts to collaboratively address social challenges that relate to their lives and communities using technology.

In sum, AEP's formative assessment and survey strategies provide instructors with data to understand students' interests, barriers to learning, and any microaggressions they face in STEM. AEP addresses cultural and social misalignments by equipping instructors with a framework to clarify expectations for success, create a space that validates students' backgrounds, and provide opportunities for students to code switch and connect their interests with academic discourse. AEP supports struggling students since it empowers learners to make meaning through productive struggle and assists students who are still acquiring the skills to complete conceptual tasks. Seeking to avoid assimilationist approaches to equity, AEP's end goal is not to ask students to think like the instructor and to solely learn a form of routine expertise via didactic modes of teaching (Schwartz et al., 2015). Instead, AEP encourages students to develop a greater sense of adaptive expertise, where students examine problems and apply relevant concepts with and without direct instruction (as described above). Throughout this process, students practice and synthesize different ways of scientific thinking gained from their community, instructors, peer-to-peer collaboration, and their own insights.

AEP is responsive to ongoing learning and is culturally responsive because it does not perceive students' pre-test scores and backgrounds as deficits dictating an endpoint. Instead, AEP sees students' backgrounds as sources for innovation, and AEP therefore empowers students to leverage their existing social and cultural competencies to apply scientific concepts to both novel problems and contexts of their choice. For example, AEP's

instructional and project-based approach enables students to co-construct and make meaning of STEM concepts to utilize innovative strategies that build on their peers' and instructors' critical thought process. Drawing on multiple frameworks, AEP focuses on equipping students with developing the mindset, skills, sense of community, and confidence to tackle meaningful problems where there exists no single solution.

## Methods

127 undergraduate student participants from a R-1 US university took the same course, but were randomly assigned into treatment or control sections. The course required students to complete a pre-test on foundational course content and a final assessment. These assessments and the following identical measures were collected in both conditions: validated surveys, demographic information, interviews, observation notes, anonymous course-feedback forms, and formative and summative assessments. Four observers took notes during every class in both conditions to determine if 1) the treatment course applied AEP and 2) the control course did not adjust instruction based on data. The treatment and the control conditions had the same instructors who employed exemplary active-learning strategies (e.g., modeling key skills, deliberate practice, dynamic lecturing, dialogue, case studies). However, instructors adjusted their teaching practices for students in the treatment group based on weekly student data. For the control condition, these same instructors did not adjust their teaching based on weekly student data, since they did not see these data. To ensure that the treatment did not impact pedagogical practice in the control condition, the control condition was taught before the treatment condition in each week of instruction. Instructors were provided with data from the treatment condition only after they had taught the control condition. In the control condition, students received scores on assessments but instructors did not.

We conducted multiple regression at the 5% significance level. We fitted one regression model for research question 1: the final assessment scores (scale 0-100) were regressed on pre-test (scale 0-100), dichotomous variables for treatment group and disability status, and a categorical variable for gender (female and male with gender non-conforming as the reference category). For question 1, the coefficient of treatment is an estimate of the difference in mean final achievement between treatment and control students after controlling for disability status, gender, and pre-test. For question 2, we looked for an interaction between treatment group and disability status within the regression model. The coefficient for this interaction term represents the mean difference in the treatment effect between disabled and non-disabled students, after controlling for pre-test and gender. We also looked for an interaction between treatment group and gender within the regression model. This produced two interaction terms. The coefficient for one interaction term represents the mean difference in the treatment effect between gender non-conforming and female gender groups, controlling for pre-test and disability status. The coefficient for the other interaction term represents the mean difference in the treatment effect between gender non-conforming and male gender groups, controlling for pre-test and disability status.

We also used Saldaña's (2009) methods of qualitative coding, categorizing, and identifying patterns. We inductively coded data by identifying themes and assigning thematic codes. We then created overarching thematic codes based on patterns, which became our analytical focus. To establish inter-rater reliability, we created a codebook with definitions of engagement and achievement with examples to guide our analyses. We triangulated data sources and multiple viewpoints (e.g., instructors, students, and researchers) to further analyze data.

## Findings

### Descriptive statistics

The treatment and control groups have similar frequencies on disability status and gender explanatory variables (See table 1). The control's mean pre-test score ( $M=29.42$ ;  $SD=12.84$ ) is comparable to the treatment's ( $M=29.70$ ;  $SD=12.24$ ). Additionally, the treatment's mean final score ( $M= 97.82$ ;  $SD=1.75$ ) is higher than the control's ( $M=83.66$ ;  $SD=8.10$ ).

Table 1: Contingency table of disability status and gender categories of control and treatment groups

|                       | Control (n=65) | Treatment (n=62) |
|-----------------------|----------------|------------------|
| Demographic Variable  | N (%)          | N (%)            |
| Disability Status     |                |                  |
| Non-Disabled Students | 42 (64.62%)    | 41 (66.13%)      |
| Disabled Students     | 23 (35.38 %)   | 21 (33.87%)      |

Gender

|                       |             |             |
|-----------------------|-------------|-------------|
| Gender Non-Conforming | 16 (24.62%) | 15 (24.19%) |
| Female                | 26 (40.00%) | 25 (40.32%) |
| Male                  | 23 (35.38%) | 22 (35.48%) |

**Regression analyses**

We conducted a regression analysis with robust standard errors to correct for heteroskedasticity, in order to examine the treatment effect on collaborative final project outcomes, controlling for disability status, pre-test achievement, and gender. See table 2 for regression results. We performed appropriate regression diagnostics to validate the use of this approach. There was no evidence of collinearity between predictors (mean VIF =1.24).

Table 2: Multiple regression with robust standard errors for the effect of treatment group, pre-test, disability, and gender

| Variable                           | Est. Coeff. (Robust Std. Err) | 95% Confidence Int.l |       | p-value |
|------------------------------------|-------------------------------|----------------------|-------|---------|
| Treatment                          | 14.20 (.99)                   | 12.23                | 16.17 | <0.001  |
| Pre-test                           | 0.03 (0.04)                   | -0.06                | 0.12  | 0.50    |
| Disability                         | 2.65 (1.16)                   | 0.35                 | 4.94  | 0.02    |
| Gender (Reference: Non-Conforming) |                               |                      |       |         |
| Female                             | -2.73 (1.40)                  | -5.50                | 0.03  | 0.05    |
| Male                               | -2.28 (1.34)                  | -4.93                | 0.37  | 0.09    |
| Intercept                          | 86.38 (2.58)                  | 81.28                | 91.48 | <0.001  |

*Research question 1.* The treatment group on average is estimated to perform 14.20 percentage points higher (SE = 0.99, 95% confidence interval from 12.23 to 16.17) on the final than the control, after controlling for disability status, gender, and pre-test achievement. This difference was statistically significant (t =14.27, df=121, p<0.001). This model’s R<sup>2</sup> is 0.63, suggesting that 63% of the variation in the final assessment scores can be explained by treatment group, disability, gender, and pre-test.

*Research question 2.* The effect of AEP persists across disability status and gender groups. No significant interaction exists between treatment group and disability status at the 5% level. Additionally, no significant interaction exists among treatment group and gender at the 5% level. Therefore, students in each of these groups appear to benefit equally from AEP.

**Qualitative findings**

At the beginning of the semester, survey data showed that treatment and control students had anxiety about collaborative learning. Many of them were unconfident in their skills and they often felt uncomfortable interacting with peers because they did not know each other or their peers’ strengths. Students also felt that project requirements were overwhelming since they had multiple parts and asked for deep conceptual application. One student stated, “Working with others on a project that has no clear answer feels much more difficult than a multiple-choice exam.” Students also added that using statistical software felt daunting and they had fears about their team not completing project tasks. To alleviate student anxiety, instructors applied AEP strategies to address interpersonal and conceptual shortcomings that students voiced on formative assessments and surveys. In surveys, treatment students indicated that it was helpful when instructors provided team activities and peer-to-peer validation exercises to address these shortcomings and foster a sense of community.

Coding of survey data suggests that AEP provided a structure that helped improve students’ collaborative learning skills and experiences. Treatment students indicated that they learned effectively from applying statistical competencies via peer-to-peer activities to succeed in a collaborative scientific community. These students noted that group members contributed effectively to the team since peers and instructors ensured that everyone understood how to apply course concepts and use software. These treatment students highlighted how instructors

supported groups by providing additional examples of how other organizations dealt with interpersonal challenges. Moreover, treatment students appreciated completing point-people, task, and deadline sheets, because they could plan and break down projects into manageable tasks that could be completed each week within an hour. These tasks had deadlines with enough time for the team to provide feedback to each other and to allow extensions that did not significantly slow down the team. A majority of students also found it less stressful when they could submit project pieces in small chunks to a folder or other medium (e.g., Canvas page, Google drive, project Dropbox folder), because they could earn points for these smaller and intermediate submissions. Treatment students also highlighted how instructors used assessment data to provide resources and page numbers from readings to address group project needs. Moreover, the teaching assistants monitored who was completing tasks on time. They provided suggestions to group members on how they could ask questions to understand and address the barriers that team members had with completing tasks on time. The teaching assistants regularly encouraged team members to play to their strengths and to offer each other collaborative support with completing tasks. Students also mentioned that ungraded individual reflections mitigated conflict and contributed to more productive collaboration, since students self-acknowledged and addressed their areas for growth. This study shows that formative assessment data in higher education is not only useful for addressing students' conceptual difficulties, but is also beneficial for addressing barriers to collaborative learning.

Based on surveys and interviews, treatment students reported excelling on the collaborative final project because 1) they received weekly feedback from peers and instructors and 2) instructors adjusted teaching in response to student performance. When asked to indicate effectiveness of various teaching practices, treatment and control students appreciated how steps for challenging problems were broken down, linked to definitions of concepts, provided in writing, and modeled for them with reference to a rubric. Treatment students reported being more prepared for collaborative in-class activities when instructors provided class-discussion questions and lesson-plan outlines before class. Treatment students also consistently emphasized how instructors addressed misconceptions from weekly assessments, reviewed challenging concepts, and validated and built on the different ways students arrived to accurate solutions. Additionally, these students reported improved learning when instructors offered written cues and strategies to help students analyze question prompts. These students said this approach helped them apply different skills and thought processes when completing assessments. Moreover, treatment students stated that the collaborative learning process was helpful because they were held accountable to a peer and had to explain and answer questions about the rationale behind their thought processes. Treatment students appreciated using various concept-driven step-by-step strategies to identify their problem and goal, select the strategies to address their goal, and then develop a rubric for assessing their work and their peers' work. These students mentioned that this process helped them engage in purposeful application and avoid mindless application of code and strategies when analyzing data.

By contrast, data from the control condition show consequences when instructors were not aware of barriers to student collaboration. Many control students mentioned experiencing difficulties with collaborative assessments since they needed more review of dense concepts and software skills to contribute effectively to their teams. Moreover, several control students indicated that they were not satisfied with group projects because collaboration was strained. For example, some control students felt that they could not voice their concerns to their instructors, fearing repercussions for their teammates. Many control students indicated that they decided to work on tasks individually to avoid conflict. Fortunately, a majority of students in both conditions found synchronous and asynchronous computer collaboration to be helpful for sharing code, editing visualizations, and analyzing data together, especially when they used Google doc functions. For example, these students used collaborative editing features and clicked a user's icon in the Google doc, which would bring them to the location within the document where their counterpart was working. Students in both conditions reported learning effectively when they collaboratively worked on clicker-esque exercises because they could explain concepts to each other and question each other's thinking.

## **Implications and conclusion**

AEP's large standardized effect size on the final ( $d=2.40$ ) is consistent with previous studies that examined the impacts of formative assessment and active learning (Deslauriers et al., 2011; Froyd, 2008; Phuong et al., 2017). AEP provides a framework on how to leverage the benefits of extended and embedded collaborative learning. Students in the treatment reported how learning to diagnose and break down problems from instructors and peers helped them better understand concepts since they could see the application of these concepts in collaborative tasks and via computer-based visualizations. Highlighting the benefits of embedded learning, treatment students emphasized that the material felt more relevant when they were asked to collaboratively discuss different step-by-step modeled strategies that were and were not useful for the context of their team project and why. Alluding to the advantages of extended learning, treatment students also said the concepts and statistical tests made more

sense and were worth learning when they were applicable to contexts and data that were meaningful. These findings suggest that AEP can help instructors address their students' learning needs (i.e., their interests, strengths, and areas for growth) to improve computer-supported collaborative learning experiences in real-world contexts.

This research is novel since no study to our knowledge has used an RCT with the same university instructors to compare an equity-oriented framework with a control condition that employed exemplary active-learning strategies. This control condition avoided the pedagogical limitations cited in multiple studies across different higher education contexts for introductory STEM courses. Pervasive across several colleges, including the university in this study, these limitations include curved grading, competitive and un-collaborative peer environments, overpacked lecture-based curricula, limited faculty interaction, and artificially difficult exams that are disconnected from both classroom instruction and the real world (Bettinger 2010; Barr, Gonzalez, & Wanat 2008; Crisp, Nora, & Taggart 2009; Eagan et al. 2011; Seymour & Hewitt 1997). At the R-1 institution we studied, statistics instructors in introductory courses and electives primarily teach through lecture and typically do not adjust teaching based on weekly formative assessments aligned with the final's rigor. They often have problem sets and classroom exercises that are lower in rigor than the final and midterms. For the active learning control condition in this study, instructors modeled key skills and provided deliberate practice and active learning activities to students. However, like many higher education STEM faculty who have not been formally trained in teaching, the control instructors did not adjust instruction based on weekly assessments. Nevertheless, control students, on average, performed about a third to half a letter grade higher than department norms for introductory statistics courses. By comparison, the treatment group, on average, scored over a full letter grade higher than these department norms. Therefore, there are significant implications for AEP in introductory college statistics and data science courses that are foundational to STEM disciplines. These courses can be gateway courses, which impact students' attitudes, confidence, sense of belonging, persistence, retention, and identity in STEM. Applying AEP can benefit universities and STEM-related departments that want to create collaborative spaces that retain URM.

Limitations to this study include institutional context and that the disability variable combines physical, invisible, and mental health disabilities. Another possible limitation is that the control was taught before the treatment for each week of instruction. However, in previous studies, the standardized effect size ( $d=2.36$ ) was similar when the AEP treatment condition was taught before the active learning control condition (Phuong et al. 2017). Future research directions include controlling for additional variables (e.g., race, class) and finding ways to support faculty with implementing AEP to improve student learning. It would be particularly interesting to examine whether students in the treatment group would perform better in subsequent courses. We would also like to examine if the treatment would impact other academic (e.g., persistence, retention) or psychosocial outcomes (e.g., sense of self-efficacy, sense of belonging, stereotype threat) in STEM courses.

Faculty development centers and equity units can provide programs, guides, and incentives that support instructors with applying elements of AEP. AEP's elements (e.g., formative assessment) can be useful across disciplines, academic departments, universities, and schools committed to narrowing academic achievement gaps, especially for URM. By increasing URM students' success in higher education, the AEP model strives to diversify perspectives on STEM and respond to systemic inequities, such as retention issues. These students' success, which can be enhanced through AEP, can inspire others in their communities to pursue STEM-related majors and careers.

## References

- Allen, R. A., Folkhard, A., Abram, B., & Lancaster, G. A. (2010). Statistics for the biological and environmental sciences: improving service teaching for postgraduates. *Journal of Statistical Education*.
- Deslauriers, L., Schelew, E., & Wieman, C. (2011). Improved learning in a large-enrollment physics class. *Science*, 332(6031), 862-864.
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23), 8410-8415.
- Hattie, J. (2009). *Visible learning: A synthesis of meta-analyses in education*. Routledge.
- McGregor, K. K., Langenfeld, N., Van Horne, S., Oleson, J., Anson, M., & Jacobson, W. (2016). The University Experiences of Students with Learning Disabilities. *Learning Disabilities Research and Practice*, 31(2), 90-102.
- Phuong, A. E., Nguyen, J. & Marie, D. (2017). Evaluating an adaptive equity-oriented pedagogy: A study of its impacts in higher education. *The Journal of Effective Teaching*. 17(2), 5-44.
- Schwartz, D. L., Tsang, J. M., & Blair, K. P. (2016). *The ABCs of how we learn: 26 scientifically proven approaches, how they work, and when to use them*. WW Norton & Company.
- Seymour, E., & Hewitt, N. (1997). *Talking About Leaving: Why Undergraduates Leave the Sciences*. Westview Press.

# The Effectiveness of Publicly vs. Privately Assigned Group Leaders Among Learners in Rural Villages in Tanzania

Judith Uchidiuno, Evelyn Yarzebinski, Emilio Vargas-Vite, Ken Koedinger, and Amy Ogan  
jio@andrew.cmu.edu, eey2@andrew.cmu.edu, evargasv@andrew.cmu.edu, koedinger@andrew.cmu.edu,  
aeo@andrew.cmu.edu  
Carnegie Mellon University

**Abstract:** Research studies show that teachers increase the success of education technologies in rural settings by supporting students via technology support, domain-relevant explanations, enforcing discipline, and maintaining student engagement. However, a teacher's presence hinders student collaboration in some cultural contexts, and some students may not have a teacher or knowledgeable adult who can provide this support. We conducted an experiment with K-1 students (N=36) in a rural Tanzanian village, where we trained students to provide technology support for their peers under different experimental conditions. We found that with basic technical training and social awareness of the assigned leaders, students can indeed provide peer support in the absence of a teacher, and additionally enable collaboration. We challenge the popularly held notion that natural leaders will emerge and support students' technology and learning needs without adequate training, and discuss the implications of our findings in the deployment of technologies in similar socio-cultural contexts.

## Introduction

The number of out-of-school children in rural and urban regions of Sub-Saharan Africa is increasing at rates that traditional schooling infrastructures cannot accommodate. Current estimates by UNESCO report that there are over 63 million out-of-school children, with over 50% of those children living in Sub-Saharan Africa (1). Those estimates have been increasing over the last few years despite government commitments to address the issue. Of the children enrolled in school, less than 20% score above minimum proficiency on math and reading assessments (2). Popular initiatives to address this challenge involve deploying educational technologies to students in and out of schools, particularly as hiring the large quantity of necessary qualified teachers is difficult (Buchele & Owusu-Aning, 2007). However, such wide-scale initiatives have mostly been unsuccessful due to a lack of understanding of the local context of the end users, as well as a lack of technical infrastructure and support for these initiatives (Therias, Bird, & Marshall, 2015; Warschauer & Ames, 2010).

Despite the failure of these large-scale, one-size-fits-all initiatives, several studies have shown that when deployed using a bottom-up approach, where the social, cultural, and infrastructural needs of the end users are prioritized, educational technologies can indeed lead to positive learning gains (Hollow & Masperi, 2009; Therias et al., 2015). These studies emphasize that providing support for students, including technical and domain knowledge support, is critical to the success of education technologies in under-resourced communities. Other types of support needed by students include help navigating learning applications, scaffolding their lack of basic digital literacy, and maintaining engagement while learning with technology (Uchidiuno et al., 2018).

A number of these large-scale education technology initiatives have been developed with the expectation that natural leaders will emerge among peers and provide this much needed support, without the need for a knowledgeable adult (Mitra et al., 2005; Papert, 2006). Some prior CSCL research has investigated the emergence of leaders among collaborating groups (Sun, Jackson, Burns, & Anderson, 2017), including peer-nomination and self-perceived leadership qualities (Xie, Sun, & Lu, 2015). In our study, we observed groups of K-1 students in a rural village in Tanzania using an Android tablet-based literacy and numeracy curriculum without the support of a teacher. As our own recent work has indicated that emergent leadership is rare in this context, in this study we build on prior work by exploring the effects of assigning leaders within a group of learners, under two conditions that differentially bestow authority on these leaders: a publicly assigned group leader; and a privately assigned group leader (in which the leader is assigned, but not publicized to the rest of the group). Based on scores from an administered pretest, we trained selected leaders on how to navigate the tablet application and perform basic troubleshooting. We investigate students' help-seeking and help-giving behaviors in these different conditions and provide insights on how to better foster peer support for educational technology in the absence of a teacher or knowledgeable adult.

## Literature review

The classroom is an organized system of social interactions with institutional and socio-cultural norms influencing students and teachers behaviors (Ädel, 2011). These norms influence help-seeking and help-giving behaviors, and may affect the efficacy of educational interventions if unaccounted for. For example, in certain contexts, teachers encourage children to collaborate freely with one another (Halloluwa, Vyas, Usoof, & Hewagamage, 2018; Mann, Hinrichs, Read, & Quigley, 2016), while in others, collaboration in the presence of a teacher is regarded as cheating (Uchidiuno et al., 2018). Peers have been found to support each other by engaging in discourse that can bring about positive learning gains, including “modelling, assisting, directing, tutoring, negotiating, affirming, and contradicting each other” (Stone & Christie, 1996). Without training in meaningful help-giving behaviors, students are likely to provide just the answers to their peers’ questions rather than elaborated, domain knowledge explanations (Roscoe & Chi, 2008). The quality of this discourse depends on students’ prior knowledge and self-efficacy, as well as their personal relationships with one another (Parr & Townsend, 2002). Prior studies also provide evidence that the quality of help peers provide for one another depends on the closeness of their friendships (Graesser & Person, 1994; Webb & Mastergeorge, 2003).

The support that students offer one another becomes even more important when educational technologies are deployed in unsupervised settings, where children have to rely on each other to navigate and learn from such systems. Some research studies have reported positive outcomes from such scenarios, reporting an increase in student curiosity (Mitra & Dangwal, 2010), and even marginal learning gains (Breazeal, Morris, Gottwald, Galyean, & Wolf, 2016), but do not provide any insights on the social interactions that lead to such outcomes. For instance, Kumar et al. provide rich insights from conducting an unsupervised learning research study using a mobile phone in rural parts of India (Kumar et al., 2010). They found that beyond infrastructural issues such as inconsistent electricity, factors such as gender, caste, and time of day significantly affected mobile usage of learning content. Children learned on average three new words per week for 9 weeks. The authors report that children learned to support their peers from watching the experimenters navigate the devices but provide few details on how this transfer of knowledge occurred.

While these studies expected (and in some cases did observe) natural leaders to emerge amongst the group, other research studies have provided insights on how peers interact when they are explicitly assigned to offer support in learning contexts. Fantuzzo et al. examined the efficacy of reciprocal tutoring, where students take turns providing support for their peers, and fixed-role tutor assignments (Fantuzzo, King, & Heller, 1992). Reciprocal tutoring was found to have more benefits over fixed-role assignments, however, is impractical for contexts such as our target demographic where most students are learning using a technology medium that is foreign to them. Group leaders tend to provide domain knowledge support in primarily two ways - providing answers only or stating facts (knowledge-telling) and giving reflective and elaborated explanations (knowledge-building). These types of support differ not only in the quality of feedback a help-seeker receives, but also benefit the help-giver differently - peer-tutors who provide knowledge-building support score higher on posttests as they organize their knowledge and monitor their understanding better (Roscoe & Chi, 2007). When students are not trained to provide constructive and knowledge-building feedback to their peers, they may likely provide knowledge-telling feedback, but studies show that deeper questions from help-seekers can elicit more elaborate, knowledge-building responses (Duran & Monereo, 2005).

Research such as (Kumar et al., 2010) has provided limited evidence that when children are equipped with technical skills, they can naturally emerge as leaders and provide support for their peers, without having an official assignment as a group leader. Other studies show that behaviors differ when leaders are assigned rather than emergent (Wickham & Walther, 2009) We investigate whether this finding applies in a different cultural context by equipping children with basic technical knowledge, and asking them to support groups where they are officially assigned as the group leader. We deployed an Android tablet-based learning system with *unsupervised* groups consisting of a child with domain knowledge and technical competency, their closest friends, and randomly assigned peers. We vary the experiment conditions by publicly assigning a leader in some groups, and not in others. We use these observations to answer the following research questions:

- To what extent do knowledgeable children take on the role of a leader within a group of learners, when either privately or publicly assigned the authority to do so?
- What kinds of support (knowledge-telling vs knowledge-building) do assigned leaders offer across social factors such as gender and close friendships in this cultural context?

## Methodology

This study was conducted in partnership with a Swahili-speaking, rural village in a Northwestern region of Tanzania. Members of our research team have conducted research in the region over the last three years. The village was limited in physical and technical infrastructure with inadequate power or clean water. Three mobile



network providers serviced the village. Throughout our stay (and in previous visits), we observed families with mostly feature phones without internet connectivity, however, a few school teachers owned basic Android smartphones. We did not observe any tablets in the village homes, schools, or public settings throughout our stay. Consent (in Swahili) and approvals were obtained from students, parents/guardians, the school administration, and the village council with help from a native Swahili speaker.

The Swahili learning application focused on the following areas: literacy (letter and phonemic awareness, writing, stories curated from the African Storybook Project - <http://www.africanstorybook.org/>), and math (number identification, number writing, addition and subtraction). Most system interactions involved tapping on the screen, although some require tracing, writing, or speaking to engage, with a speech recognition engine validated in various African contexts (Mills-Tetty et al., 2009), and video tutorials with continuous finger placement scaffolds to support children's' digital literacy . The Swahili audio prompts and instruction were recorded by a Kenyan Professor of Swahili, who has taught Swahili from kindergarten to university levels. The Swahili video application tutorials were recorded by a Tanzanian instructor of Swahili, who grew up in a region about 10 hours away from the village where the study was conducted. All communications with the children were translated by a native Swahili speaker who lived in the same village and was well-known to the children. More details about the learning application can be found in (Uchidiuno et al., 2018).

To ensure that we observed a sufficiently broad sample across experimental conditions, we structured our observations into 24 discrete sessions during which six groups of six children each engaged with the tablet software. Each observation lasted about 1 hour. In these sessions we observed a total of 36 unique children; this included 18 girls and 18 boys from grades K-1, ages 5 to 10 depending on year of entering school. Before the sessions began, we gave all the children a paper pre-test covering letter and number recognition, letter and number writing, simple word problems, and listening comprehension. These pre-tests were administered by Swahili speakers who read the questions one by one to each child, and recorded their responses. We noted down the students who scored the highest, and distributed them equally among the groups. Following the pretest, we ran a baseline session with each group, allowing students' natural interaction patterns to emerge. We selected six children who performed well on the pre-test *and* quickly learned to navigate the tablet application without adult assistance. We selected the highest-scoring children as leaders just in case their peers needed domain knowledge support in addition to technical support, and assigned them as leaders for each group in the experiment scenarios. We started forming the groups by asking each leader to select two of their best friends from the class - each group was comprised of a leader, the leader's selection of their two best friends, and a random assignment of three other children from the class, balancing each group by gender. After forming groups, we informed the leaders privately that they performed the best in the pre-test and with navigating the system, and were now responsible for supporting their group. We conducted a training session with all six group leaders to reinforce practices of navigating each application without assistance, as well as performing basic troubleshooting tasks e.g. helping a peer return to the learning application if they exit accidentally. After the training, all leaders could navigate the tablet and application without assistance.

Table 1: Group Information with Summary of Pre-Test Scores, the leader's Score, and Experiment Condition

| Group # | Age Range | Score Mean | Score SD | Score Range | Leader's Score | Leader Condition          |
|---------|-----------|------------|----------|-------------|----------------|---------------------------|
| 1       | 5-10      | 34%        | 23%      | 8-72%       | 72%            | Privately Assigned Leader |
| 2       | 5-7       | 46%        | 14%      | 42-60%      | 60%            | Privately Assigned Leader |
| 3       | 5-7       | 46%        | 12%      | 32-62%      | 62%            | Privately Assigned Leader |
| 4       | 6-9       | 72%        | 14%      | 52 - 90%    | 90%            | Publicly Assigned Leader  |
| 5       | 6-8       | 79%        | 11%      | 65-93%      | 90%            | Publicly Assigned Leader  |
| 6       | 6-8       | 71%        | 13%      | 55-88%      | 88%            | Publicly Assigned Leader  |

Before the start of the sessions, we reminded the leader of their job to help their peers, as no adults would be around. For 3 groups, we shared this privately with the leader so other group members did not know the group had a leader (privately assigned leader condition), however, in the other 3 groups, we made this announcement publicly, informing the group that the leader would answer all questions they had (publicly assigned leader condition). All groups completed 4 sessions each - 1 initial baseline session without a leader assigned and 3 sessions with an assigned leader. Following the baseline, we conducted two periods of classroom observation to better contextualize each leader's natural peer interactions. leaders are summarized below as "L#"; the number corresponds with their group:

- L1: 6 y/o girl - Quiet. Answers all questions correctly, does not interact much with her peers.

- L2: 7 y/o boy - Energetic, sometimes disruptive. Tries to answer all questions asked by teacher before others can, so the teacher sits him in the back, and never calls on him to answer questions. Once when the teacher left, he ran to the front and pretended to teach the classroom.
- L3: 6 y/o girl - Usually answers correctly when called on. Observed mouthing an answer to a boy called to answer question when he did not know the answer. Usually sits with her friends in the front, who all wait for her to write an answer on a worksheet, and then then copy her for in-class activities.
- L4: 9 y/o boy - Quiet in class. Does not speak except when directly asked a question, does not disrupt.
- L5: 8 y/o girl - Quiet but engages in occasional banter with friends even when the teacher is present.
- L6: 7 y/o girl - Quiet but also engages in occasional banter with her friends.

No researcher or adult was present in the room with the students during the study sessions. Instead, these unsupervised sessions were video recorded from multiple angles to capture natural student interactions. The team analyzed the data from all group sessions, recording qualitative observations of the student interactions with the tablet and each other. Next, the team reviewed the videos, notes, and observation logs and identified emergent themes related to the types of help-seeking and help-giving behaviors surrounding each group leader following a grounded theory approach (Corbin & Strauss, 2008). Six themes of student behaviors emerged: 1) student distracting the group, 2) leader addresses student distraction, 3) student asks leader for help, 4) student asks leader to help a peer, 5) student (not leader) helps another non-adjacent student, and 6) leader helps a non-adjacent peer. Next, we split the sessions equally among 2 team members for categorization according to the 6 themes. A third team member reviewed and coded all 24 sessions to validate the categories created by the other team members, breaking each session into 30 second intervals and adopting a partial-interval recording method (Hintze, Volpe, & Shapiro, 2002). Finally, we triangulated all observations related to these themes with logs captured by researchers in the field, debrief recordings, follow-up interviews with teachers, and photographs captured on site to ensure that all evidence were mutually supportive. After all sessions were categorized, all members of the team each reviewed the findings for all 24 sessions, discussed all areas of disagreement, and re-categorized findings as agreed upon by the entire team. We worked with native Swahili speakers from Tanzania to help translate interactions, as well as provide insights on the cultural underpinnings of those interactions. This study design and data analysis methodology has been used and validated in previous CSCL research studies e.g. (Pauw et al., 2015).

## Findings

### Baseline student behavior

As soon as the sessions started and the adults left the room, the children smiled at each other, looked around briefly, and then started interacting without prompting. These interactions included sharing new activity types with their adjacent peers, sharing a funny activity, celebrating their accomplishments, singing nursery rhymes together, and even repeating spoken tablet instructions in unison. We observed lots of play (and real) fighting in the room as they encroached on each other's tablets, and children intentionally distracting each other from staying on task. If an adult re-entered the room for any reason (forgot an item, child in the session called for their help, end of session, etc), the children stopped all interactions and focused on their tablets.

Without an assigned leader, students generally provided help when **all** of the following conditions were true: 1) they noticed a peer struggling or the peer asked them for help, 2) the peer is sitting adjacent to them, 3) they are not engrossed in their own work, and 4) they knew how to help (vs trying to find a solution to a new problem). In general, their behavior mirrored the classroom observations - quiet students kept to themselves even when other children tried to interact with them uninvited, and talkative students were the most vocal (and disruptive) in their sessions. The students connected the learning application to what they learned in the classrooms - on several occasions, we observed children bringing out their notebooks from their bags to double check answers before they selected answers on the tablet. They primarily communicated with their adjacent peers, and those whose screens they could easily see. Across all six groups, we observed only two instances where a student helped a non-adjacent peer. Figure 1 shows the student seating arrangement in the classroom. Students sat in a row, all facing the same direction, allowing them to interact with their peers and see their work. There were no observable differences in how girls interacted with other girls vs boys (and vice versa), although friends seemed more likely to share funny stories and new activities with each other.

In all interactions observed under this condition, students provided knowledge-telling support to their peers either voluntarily or when solicited; they either selected the answers for their adjacent peers or told them what answer to tap without any elaboration. We observed one instance of student collaboration in this condition

(and another in the publicly assigned leader condition) when knowledge-telling was insufficient. In one interaction with L5, an adjacent peer was struggling with a tracing task. After she asked L5 for help, L5 said “andika” [Swahili for “write”]. The girl followed L5’s instruction, but her answer was still not accepted because she wrote beside (rather than over) the trace outline. Then L5 reached in to help but also wrote beside the outline, resulting in a rejected response. They looked for a teacher, but no one was close by, so they tried different strategies until they learn the correct way to trace on the tablet.

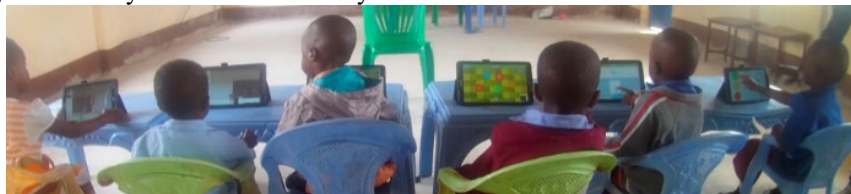


Figure 1. Typical seating arrangement in experiment sessions.

### Student behavior in privately assigned leader condition

Students in this condition behaved similarly to the baseline condition. L1 remained quiet throughout her sessions, only interacting with adjacent peers, and volunteered help when she noticed a struggling adjacent peer. Since she was not revealed as the group leader, non-adjacent students did not ask her for help. When engrossed in her tablet, she ignored most requests for help even from adjacent students. This observation was consistent in all of L3’s sessions as well. She freely chatted with adjacent peers, but also ignored their requests for help when engrossed in her own work. On the occasions where L1 and L3 helped their adjacent peers, *they showed very little persistence* and routinely abandoned them if their attempt was not a quick fix on the first or second try.

There were clear negative consequences when a student needed help but didn’t receive it. In Group 1, a boy struggled with basic application navigation, and could hardly complete a single learning activity in all three sessions despite sitting next to L1 in one session. Also in a session with L1, a student gave up and left the session after trying for 30 minutes to start a learning activity (including asking her for help twice), while another started to wipe tears from her eyes after repeated unsuccessful attempts at getting herself back on task. Figure 2 shows a boy in Group 3 who needed help, and his reaction when he finally learns to navigate the application after receiving help from the girl right beside him (not L3).

L2’s behavior as a privately assigned leader was quite different from L1 and L3. In the first session, he interacted mostly with one of his best friends similar to his behavior in the baseline condition. After privately reminding him of his responsibilities prior to the next session, he spent most of his time walking around the room and helping every struggling student (similar to how his teacher behaved in his classroom), without paying much attention to his own tablet. He was only asked for help 4 times from his adjacent friends throughout the session, but volunteered help to 18 additional non-adjacent peers. When his peers noticed that he was providing this help, they called on him to help other struggling peers, and he continued to hover around their tablets to review their progress. If a peer paused briefly, he ran over to solve the problem even if they did not ask for help. Some did not welcome the constant, unsolicited help - near the end of one session where L2 “helped” a girl constantly, she shoved him away when he walked over to her. His monitoring also quickly turned into behavior enforcement. One of the activities in the learning application involved a racecar that made screeching sounds. Students quickly figured out how to exploit this sound and enjoyed producing an almost-constant screech, therefore distracting the whole group. In the first session, L2 and his adjacent peer engaged in this distraction, but by the second session, he monitored and turned off students’ tablets if they tried the exploit. His enforcement tactics became progressively stronger in each session, and he started hitting kids on the head and addressing the whole group sternly when he thought they were not paying attention. Although some children had problems with his style of help-giving, overall it had positive effects on the group. Most children stayed on task, students started calling him for help for themselves and for others, and others started to emulate him by standing up and walking around to help as well. Figure 3 shows L2 addressing a student who was distracting the group. On the average, L1 and L3 ignored 79% of help requests overall (and all requests from their friends), while L2 ignored only 33% of his requests for help (13% from his friends).



Figure 2 (L-R). Struggling boy; peer notices and helps; excitement when he gets questions correct.



Figure 3. L2 addressing a student who was distracting the entire group.

### Student behavior in publicly assigned condition

Publicly assigned leaders had fewer help-giving restrictions; they generally provided help either if they were asked or if they noticed a peer struggling. Unlike L2, none of the leaders in this condition (L4, L5, L6) hovered around the entire group. Despite their working steadily and constantly on their tablets, they responded to most requests for help. The other children in the groups chatted with their adjacent peers about new activities and funny stories, and even attempted to help each other first before calling on the leader. On the average, leaders in this condition responded to requests for help 75% of the time (77% from their friends). They were able to discern when they were called for non-help requests such as discovering new application activities. L5 was even able to tell when she placed someone in an application that was too easy for them. She restarted a boy's application, and placed him in an alphabet song. Immediately, she exited the application, and switched it to a more difficult story for him to read. Although the leaders were called upon frequently in the first session, the help requests decreased in subsequent sessions, suggesting that students became more proficient at navigating the tablets without the leader. By the 3rd session, the publicly assigned leaders got the same number of requests from the whole group as the privately assigned leaders from their adjacent peers. Figure 4 shows a chart of the average number of times the leaders were called upon as the sessions progressed per condition.

The biggest difference between the help offered by publicly vs privately assigned leaders was the degree of leader persistence. We did not observe any cases where the leader abandoned a student even when they could not figure out a solution. One student in Group 4 accidentally exited the application but could not open a new instance. She asked L4 for help, but he could not figure out a solution after several attempts. He started to walk back to his seat but changed his mind, knelt beside her, and kept trying. After much trial and error, she found the application switcher button and selected the learning application. Instead of returning to his seat after this success, he reproduced the problem, so he could practice the solution she discovered. Following that, he returned to his seat. Similarly, L5 spent over 5 minutes trying to help a girl with a technical issue with constant application crashes. After several unsuccessful tries, L5 left the session to seek a help. The application had to be reinstalled to fix the error.

Finally, the leaders helped to maintain engagement with the learning applications. All students enjoyed exploiting the racecar game's screeches. Unfortunately, the noise distracted everyone in the group and increased off-task behavior. It caused students who were trying hard to focus to start tapping the tablet without regard for the current activity, or scribbling furiously in writing activities. Unlike the privately assigned leader condition, the publicly assigned leaders initially joined the fun, but quickly worked to establish order in the groups. L4 warned the students to stop misusing the application; L5 and L6 resorted to turning off students' tablets if it distracted their group, and in one incident, L6 even seized a peer's tablet. Other members of the group began to call for the leader when someone was bothering them individually, and the leaders routinely intervened. This was an unexpected but welcome benefit of having publicly assigned leaders in the group because these interventions minimized the time that students were off task compared to the other leadership conditions.

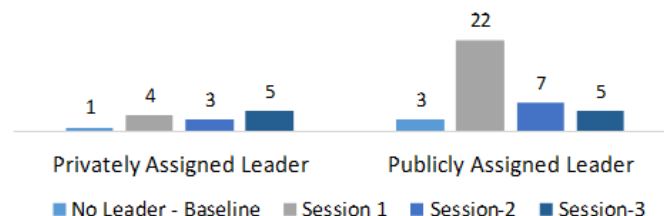


Figure 4. Average Help-Seeking Requests to in Privately vs Publicly Assigned Leader Conditions.

### Discussion and conclusions

Unassisted learning educational settings are becoming more popular due to the increasing demand for education where traditional schooling facilities and instructors are scarce - our research contributes to the body of

knowledge that maximizes their success. The expectation that children, without formal training or assignment, will provide support in unassisted learning situations seems intuitive given their innate curiosity, and previous findings from other cultural contexts. However, our research provides evidence that socio-cultural factors do not only affect the ways that children engage with educational technologies, but also in their willingness and ability to support one another, as well as the quality of help they provide. This is not to say that natural leaders in this cultural context do not exist - a student like L2 demonstrated his enjoyment for teaching and proctoring both in the classroom, and in the experiment sessions. However, in a culture like the one we studied where teachers and adults are likely the singular source of authority, this kind of behavior may be punished similar to seating L2 at the back of the class, and refusing to call on him or answer any of his questions. We also provide evidence that children who show natural helping tendencies in the classroom (like L3 whose classmates depend on her for answers), even when trained, do not automatically emerge as leaders for their peers at the expense of their own interest. Our research shows that in this cultural context (and similar), help-giving did not vary by friendships or even gender, and students provided the best quality of support when they were trained to help **and** when there was a social expectation of them providing such support. Leaders did not just naturally emerge similar to other studies exploring unassisted learning situations e.g. (Kumar et al., 2010) and (Mitra et al., 2005). Rather than expecting children to naturally provide support for their peers, such unassisted learning programs should train and educate children on who, how, and when to ask for support to maximize their success.

Publicly assigning the role of a leader conferred significant benefits to all students (leaders included). Group members, regardless of their proximity to the group leader received much needed help with the learning application, and we observed no incidents of children frustrated, abandoning the sessions, or going multiple sessions without engaging with the learning application (unlike the privately assigned leader condition). Leaders showed that they could regulate their learning and that of their peers, deciphering when a student really needed help vs when they could figure it out for themselves, and leaders became more persistent and provided better support by practicing more nuanced ways to bail their peers out of trouble, improving the collective knowledge of the entire group. These results are in accordance to previous findings that show that students' ability to influence and adjust their own cognitive and help-giving behaviors and that of others is optimal for learning and working together (Zimmerman & Schunk, 2011). Another unexpected but very positive effect of publicly assigning leaders was how much the leader helped to control disruptions to the group. In our experiment sessions, when one student started causing disruptions, the whole group (including those who were trying hard to stay focused) got distracted as well. Having students who automatically took the responsibility upon themselves to manage disruptions was important for maintaining student engagement. Finally, our research provides evidence that in a cultural context where collaboration is often discouraged by teachers e.g. (Uchidiuno et al., 2018), carefully designing unassisted learning environments can foster collaborative behaviors even for simple writing tasks. Further research is required to investigate how these behaviors change as the tasks become more complex, or when the students are explicitly engaged in collaborative tasks.

Our study is limited in several ways - the lack of in-depth explanations provided by students may be a factor of their age and developmental level, as well as the domain area not requiring much explanation (early math, letter identification, and early reading). Future studies should investigate these help-seeking and help-giving behaviors among older students in domain areas that require more explanations such as math and science targeted at later grades. Also, the only training children received was related to technology support. They were not trained on how to properly give domain knowledge feedback to their peers. Although, our study provides evidence that untrained students do not provide knowledge-building naturally, further investigation is required to determine the help-giving behaviors of students who are trained to give proper feedback. Finally, in our study we qualitatively observed a small group of high performing leaders across multiple sessions - these insights may be different in a larger scale study, or with students with lower achievement.

## Endnotes

- (1) <http://uis.unesco.org/sites/default/files/documents/fs48-one-five-children-adolescents-youth-out-school-2018-en.pdf>
- (2) <https://www.brookings.edu/blog/africa-in-focus/2017/10/06/figures-of-the-week-africa-education-world-development-report-2018/>

## References

- Ädel, A. (2011). Rapport building in student group work. *Journal of Pragmatics*, 43(12), 2932–2947.
- Breazeal, C., Morris, R., Gottwald, S., Galyean, T., & Wolf, M. (2016). Mobile devices for early literacy intervention and research with global reach. In *Proceedings of the Third (2016) ACM Conference on Learning@ Scale* (pp. 11–20). ACM.
- Buchele, S. F., & Owusu-Aning, R. (2007). The one laptop per child (OLPC) project and its applicability to Ghana. In

- Proceedings of the 2007 international conference on adaptive science and technology* (pp. 113–118).
- Corbin, J., & Strauss, A. (2008). Basics of qualitative research: Techniques and procedures for developing grounded theory. *Thousand Oaks*.
- Duran, D., & Monereo, C. (2005). Styles and sequences of cooperative interaction in fixed and reciprocal peer tutoring. *Learning and Instruction, 15*(3), 179–199.
- Fantuzzo, J. W., King, J. A., & Heller, L. R. (1992). Effects of reciprocal peer tutoring on mathematics and school adjustment: A component analysis. *Journal of Educational Psychology, 84*(3), 331.
- Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal, 31*(1), 104–137.
- Halloluwa, T., Vyas, D., Usoof, H., & Hewagamage, K. P. (2018). Gamification for development: a case of collaborative learning in Sri Lankan primary schools. *Personal and Ubiquitous Computing, 22*(2), 391–407.
- Hintze, J. M., Volpe, R. J., & Shapiro, E. S. (2002). Best practices in the systematic direct observation of student behavior. *Best Practices in School Psychology, 4*, 993–1006.
- Hollow, D., & Masperi, P. (2009). An evaluation of the use of ICT within primary education in Malawi. In *ICTD* (pp. 27–34). Citeseer.
- Kumar, A., Tewari, A., Shroff, G., Chittamuru, D., Kam, M., & Canny, J. (2010). An exploratory study of unsupervised mobile learning in rural India. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 743–752). ACM.
- Mann, A.-M., Hinrichs, U., Read, J. C., & Quigley, A. (2016). Facilitator, functionary, friend or foe?: Studying the role of iPads within learning activities across a school year. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 1833–1845). ACM.
- Mills-Tetty, G. A., Mostow, J., Dias, M. B., Sweet, T. M., Belousov, S. M., Dias, M. F., & Gong, H. (2009). Improving child literacy in Africa: experiments with an automated reading tutor. In *Information and Communication Technologies and Development (ICTD), 2009 International Conference on* (pp. 129–138). IEEE.
- Mitra, S., & Dangwal, R. (2010). Limits to self-organising systems of learning—the Kalikuppam experiment. *British Journal of Educational Technology, 41*(5), 672–688.
- Mitra, S., Dangwal, R., Chatterjee, S., Jha, S., Bisht, R. S., & Kapur, P. (2005). Acquisition of computing literacy on shared public computers: Children and the "hole in the wall". *Australasian Journal of Educational Technology, 21*(3).
- Papert, S. (2006). Digital development: How the \$100 laptop could change education. *USINFO Webchat, via OLPC Talks*.
- Parr, J. M., & Townsend, M. A. (2002). Environments, processes, and mechanisms in peer learning. *International Journal of Educational Research, 37*(5), 403–423.
- Pauw, D., Clegg, T., Ahn, J., Bonsignore, E., Yip, J. C., & Uchidiuno, J. (2015). Navigating connected inquiry learning with ScienceKit. International Society of the Learning Sciences, Inc.[ISLS].
- Roscoe, R. D., & Chi, M. T. (2007). Understanding tutor learning: Knowledge-building and knowledge-telling in peer tutors' explanations and questions. *Review of Educational Research, 77*(4), 534–574.
- Roscoe, R. D., & Chi, M. T. (2008). Tutor learning: The role of explaining and responding to questions. *Instructional Science, 36*(4), 321–350.
- Stone, S. J., & Christie, J. F. (1996). Collaborative literacy learning during sociodramatic play in a multiage (K-2) primary classroom. *Journal of Research in Childhood Education, 10*(2), 123–133.
- Sun, J., Jackson, J., Burns, M., & Anderson, R. C. (2017). Children's Emergent Leadership and Relational Thinking in Collaborative Learning. Philadelphia, PA: International Society of the Learning Sciences.
- Therias, E., Bird, J., & Marshall, P. (2015). Más Tecnología, Más Cambio?: Investigating an Educational Technology Project in Rural Peru. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 447–456). ACM.
- Uchidiuno, J., Yarzebinski, E., Madaio, M., Maheshwari, N., Koedinger, K., & Ogan, A. (2018). Designing Appropriate Learning Technologies for School vs Home Settings in Tanzanian Rural Villages. In *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies* (p. 9). ACM.
- Warschauer, M., & Ames, M. (2010). Can One Laptop per Child save the world's poor? *Journal of International Affairs, 33*–51.
- Webb, N. M., & Mastergeorge, A. (2003). Promoting effective helping behavior in peer-directed groups. *International Journal of Educational Research, 39*(1–2), 73–97.
- Wickham, K. R., & Walther, J. B. (2009). Perceived Behaviors of Emergent and Assigned Leaders in Virtual Groups.
- Xie, K., Sun, Z., & Lu, L. (2015). The Role of Time, Engagement, and Self-Perceived Leadership on Peer-Nominated Emergent Leadership in Small Group Online Collaborative Learning. International Society of the Learning Sciences, Inc.[ISLS].
- Zimmerman, B. J., & Schunk, D. H. (2011). Self-regulated learning and performance. *Handbook of Self-Regulation of Learning and Performance, 1*–12.

# Taking the Patch Perspective: A Comparative Analysis of a Patch Based Participatory Simulation

Lauren Vogelstein, Vanderbilt University, lauren.e.vogelstein@vanderbilt.edu  
Corey Brady, Vanderbilt University, corey.brady@vanderbilt.edu

**Abstract:** In this paper we explore the implementations of an activity designed to engage students in an embodied simulation of an artistically expressive Agent-based modeling environment. We compare how two facilitators implemented the same activity, analyzing how they co-developed differing substrates with their students. We argue that instructors facilitated different embodied agent-based perspectives, which led students to develop considerably different conceptualizations of how these agents interact and how to code in this environment.

## Introduction

Agent-based modeling (ABM) offers a powerful representational infrastructure for simulating and understanding emergent phenomena and complex systems. However, it *also* provides an expressive medium for creating computational art of various kinds. In the constructionist tradition, a single movable agent (turtle) has been used in environments such as TurtleArt (Bontá, Papert, & Silverman, 2010) and more recently Scratch (Resnick et al, 2009) as a means of exploring artistic expressivity focused on a single agent. In this paper, we extend this line of work into a multi-agent environment, NetLogo (Wilensky, 1999), focusing initially on the expressive visual potential of a large grid of immobile agents (patches), which can be thought of in terms of the panels of a quilt or the pixels of an image. As computational agents, patches have Cartesian coordinates, can change color, and can hold variables. In using the NetLogo patch grid as a medium for visual art, we have found students encounter conceptual challenges and opportunities in learning to use the environment expressively. In some senses, these challenges mirror those found with artistic uses of single and multi-turtle programming, where physical, embodied simulation have provided critical entry points for learners (e.g., “playing turtle” (Papert, 1980) for single turtle concepts, and engaging in “star people” activities (Resnick & Wilensky, 1998) for multiple turtle concepts). We hypothesized, however, that learning to conceptualize patches would require different mappings from social structures to computational phenomena. Engaging in design-based research to explore this area, we were interested in how a group of learners new to ABM could use coordinated social enactment to build familiarity with multi-agent patch programming as an expressive medium. We developed activities that invited participants to physically embody patches as an assembly of social beings that can each interpret and respond to code. We designed activities to foster understandings of how these agents function computationally and what visual effects they can produce. However, we found that teacher facilitators implemented these activities in ways that promoted different conceptions of patch agents, even when these activities were highly specified in professional development and curricular materials. In our study of two cases of implementation these differences appeared to be consequential, opening up two distinct ways of using patches, both of which were powerful and expressive but in different ways. In particular, we found that taking slightly different embodied perspectives as patches could contribute to the development of considerably different conceptualizations of how these agents interact and how one codes in this environment. Thus in this paper we explore the following research question: How do facilitation choices in these socially embodied simulations shape the image that emerges for participants of how one programs in a multi-agent environment?

## Study overview

In this paper we report on data collected from the Computational-Thinking And Mathematics Play Spaces (CAMPS) project, a design based research project created specifically to look at how computational thinking and mathematics could be explored simultaneously with middle school learners in an expressive, artistic ABM environment. As the acronym suggests, this project took the form of a summer camp. The first design iteration consisted of a one-week (five-day) free summer camp for middle school students, held in a middle school in a southeastern U.S. city. The camp was titled “Code Your Art,” and was advertised as a camp involving computer programming and art for rising 7<sup>th</sup> and 8<sup>th</sup> graders (although a few rising 6<sup>th</sup> graders asked to participate and attended). Code Your Art camp consisted of two groups each with 16 students. Each group was co-facilitated by two middle school mathematics teachers; Tracy and Isabel were in room 1 and Kiara and Neil were in room 2. Six researchers and two undergraduate computer science students rotated between the rooms to collect data and provide technical and facilitation support as needed.

Over the course of the five-day camp, students created digital works of art in NetLogo to present and



share on the final day in each class' art gallery. We conceptualized using NetLogo to create digital art by thinking of pixels as computational agents (patches). To do this, we had students import images from the web into NetLogo so that they could use this computational environment to manipulate different parts of the image, changing pixels based on color, location, or both. Since we conceptualized images as being composed of computational agents (pixels) it made sense for us to introduce NetLogo from the patch perspective, having students' first introduction to NetLogo be in the context of what a single pixel/patch would interpret in the context of creating a larger composite image. Thus students' first introduction to NetLogo was not on a computer, it was in a series of participatory simulations we referred to as the Stadium Card Activities. We took inspiration from a sports arena phenomenon in which spectators are given one or more colored squares of paper. When held up in a coordinated manner the entire stadium creates a composite image (Figure 1).



Figure 1. Stadium cards in increasing scale from left to right.

Since we were neither in a stadium nor in the digital world of NetLogo, but rather in middle school classrooms, we created our own physical computational environment by tapping into the cultural syntonicity (Papert, 1980) between stadium cards in the arena and patches in NetLogo. We taped out a 2 x 8 grid on the floor (one square unit for each participant/patch) and instead of giving participants a single colored card, like in the arena events, we gave each participant five 15" x 15" colored squares (one each of black, red, orange, green, and blue), each labeled with a number on one side (0, 1, 2, 3, 4, respectively), and attached by a binder ring so that students could flip through all five colored squares. Although the numbers and colors did not correspond to the same color system in NetLogo (e.g. in NetLogo red = 15 not 1), they still allowed participants to respond to code that referred to colors either by name or number, and produced an opportunity to think about modular arithmetic (e.g. what happens when you are asked to increase your color by 7).

In the context of the camp, a series of three Stadium Card Activities were used each day to introduce new computational concepts in NetLogo: patch color, patch location, and turtle movement. For this paper we will focus our analysis on the first activity, which we used to introduce NetLogo on day one of the Camp. This activity focused on patches (with color as their primary property), and basic syntax and computational concepts (e.g. brackets, "ask patches," if/then, colors named and numbered, loops, and randomness). This activity was co-designed with the instructors who taught the camp, using pilot enactments during four days of professional development to change the original researcher-designed activity. Through these conversations with instructors two consequential changes were made. (1) Someone embodied the Observer. We even dedicated a special location in the room for the Observer to stand and give commands. This created a "Simon Says" feel to the activity that the instructors thought would be something students could easily understand, tapping into the ego syntonicity (Papert, 1980) between the agents (patches) as beings that can interpret commands and make decisions and the participants themselves who have these human abilities. The instructors suggested that this would also support students in eventually becoming the Observer themselves. (2) A slide deck with all of the agreed upon NetLogo commands for instructors to read as the Observer, along with a translation into English on each slide (e.g. "ask patches [if pcolor = black [set pcolor red]]" and "Ask all patches, 'if your color is black set it to red'"). This change allowed instructors to make connections for themselves about NetLogo syntax, such as relating brackets to quotation marks when "talking" to patches as the Observer. By the end of the professional development days, a Google slide deck was created for the instructors to use on the first day of camp. Although both teams of instructors had access to the same professional development and the same slide decks, when they ran this activity during the camp they made some changes. Although kids in both rooms playfully engaged in this social ABM environment through their embodiment of agents within the system, the changes instructors made resulted in significant differences in students' understandings of how to use NetLogo.

## Theoretical framework

As seen in the two groups during Code Your Art camp, when groups work together, they develop their own micro-cultures that shape their understandings as they learn together (Fine, 1979). This development can be traced through an analysis of co-operative action (Goodwin, 2018). Goodwin describes human interaction as a process in which participants draw upon interactional components from a "substrate" of their shared history and reuse them with modification, transforming these resources in ways afforded by the semiotic system to which



they belong so that they serve emergent interactional intentions. Substrates consist of shared resources for interaction that change as time unfolds and interactions continue. Goodwin defines a substrate as follows:

It is visibly a form of action that organizes simultaneously 1) the *past* as something relevant to the present by not only incorporating, but transforming, the materials it emerged from; and 2) the *future* that will immediately follow from it by providing a constrained but open-ended framework for subsequent action (Goodwin, p. 32, 2018).

This analytic construct makes shared interactional resources visible as influencing future interaction by emphasizing how substrates evolve over time through interaction. Components of a substrate can include discourse, material tools, gestures, the physical environment, and anything else that can ground human interaction. An analysis of the substrates developed in the enactment of the Stadium Card Activities focuses on the social nature of the design through the collective embodiment of an ABM environment. During the Stadium Card Activities, relational metaphors, students position on the grid, and the order in which code was translated were the three primary components of each room's substrates that differed, constraining participants' respective open-ended possibilities for interacting with NetLogo as their computational understandings began to develop. In analyzing the different pedagogical choices instructors made when enacting this activity, an analysis of these substrates allows us to begin to understand how different choices constrained the possibilities for interaction.

## Methods

In both rooms, the first Stadium Card Activity lasted about 40 minutes, starting with (1) a discussion about stadium cards as real life examples, (2) going through a sequence of commands with an instructor as the Observer, and (3) ending with students taking on the role of the Observer. The Stadium Card Activities were video recorded using four cameras situated around the periphery of the classrooms. Researchers observed this activity and took notes while watching. In both rooms the second author walked in during the activity and participated as well. This paper reports initial findings from our review and analysis of recordings of the first Stadium Card Activity. Using Interaction Analysis (Hall & Stevens, 2016; Jordan & Henderson, 1995) we have developed grounded theoretical categories to describe similarities and differences in the substrates (Goodwin, 2018) present in each room, looking specifically at facilitation and pedagogical choices instructors made and how they related to how students made sense of new computational concepts in this socially embodied ABM environment as they began to explore the expressive potential of the system.

## Analysis

The Stadium Card Activity provided students with an accessible entry into an ABM environment that invited expressive play and allowed for participants to productively grapple with computational concepts. Although all four instructors participated in same pilot enactments of the Stadium Card Activities and had access to the same slide deck, both pairs of instructors made changes to the activities when facilitating them during the Camp. Three primary instructional choices created significant differences in how the first Stadium Card Activity was enacted in each room: (1) how the Observer was described and positioned, (2) where students stood on the grid, and (3) the order in which NetLogo code and English interpretation were translated. These differences affected the development of the substrate for the activity and thus how students interacted with each other, the materials in the room, and the instructors, and ultimately how students began to understand what it meant to use NetLogo as evident in how students took up the role of the Observer at the end of the activity through their participation in this socially, embodied ABM environment as they explored visual effects within this computational system.

## Three instructional choices that created substrates for student learning

### Establishing the Observer-patch relationship

Embodying agents in an ABM environment provided participants with easily understandable relational metaphors for the Observer-patch relationship. Facilitators' descriptions of their role as the Observer cued different relational metaphors, emphasizing different aspects of this computational environment. While facilitators in both rooms began by equating patches to pixels in an image, they positioned the Observer-patch relationship differently. In room 1, Tracy introduced the Observer as having authoritarian rule over patches:

One of us [instructors] is gonna be the Observer. So right now this is gonna be the Observer box. I as the Observer have the privilege of being able to change you as patches. Ok? So I can speak to you, give you a command, and then you HAVE to follow my command.

This stance positioned the Observer as being in control over patches, who did not have a choice in their response. In response to this utterance, the students jokingly groaned and laughed and Tracy quickly chuckled as well. The groans and laughter indicated how Tracy's positioning of the Observer-patch relationship was an exaggerated or satirical depiction of the teacher-student relationship, a relationship the students were all too familiar with. In this moment, Tracy explained how she planned to embody a figure with complete control over these students: what they could do, when they could do it, and how they could do it. Here, the computational agents (patches) were depicted as subservient to an all-powerful ruler (Observer). Computationally one might argue that this is an accurate depiction because unlike humans, computational agents cannot think or choose or express agency, they can and must perform the actions commanded by the code they receive.

In room 2, Kiara positioned the Observer as a messenger to the patches, communicating the request (code) of another being, who could be anyone:

And so then, I'll be the Observer and I'll be standing in the Observer box and later on you could be the Observer. And so what it's asking me, it says, "ask the patches," that's you guys. In a moment I'm gonna ask you to do something. So we always, the Observer always tells the patches what they're supposed to do.

Kiara began this introduction by telling the students that although she would play the part of the Observer at the beginning, they would get a chance to play the part later. This pedagogical move emphasized accessibility of the role as something students would be able to succeed at during the activity. The transient embodiment of the Observer also distanced the role from the person, describing a different power dynamic than Tracy constructed. Distancing the person from the role was also evident when Kiara said, "And so what it's asking me, it says, 'ask patches,'" here Kiara depicted the Observer as a messenger, relaying a message (asking patches) for someone else. Computationally, this foregrounded the role of the coder as separate from a computer/Observer that relayed code to computational agents, placing agency in the coder and not the Observer.

By positioning the Observer differently in each room, the instructors contributed to the creation of different substrates in their respective rooms, drawing on two different relational metaphors—rulers and messengers—that turned out to influence students' own relationships with this computational structure and thus their understandings of how to interact in this environment and imagine differing possibilities.

### Positions and perspectives on the grid

Physically standing in for patches on the grid provided students with a point of accessibility into the role and perspectives of this agent. The students in each room stood in different places on the grid (Figure 2), which provided students with different agent-based perspectives. In room 1, students stood inside the grid, such that each student was a patch and held their colored squares for the entire duration of the activity. This meant that these students experienced "the patch perspective" in which they could not see the entire image created by the group, only their neighbors were directly visible. In room 2, students stood outside of the grid facing inwards, meaning that unlike a digital patch, students could see the entire image created by the group. This difference meant students' perspectives on and visual access to the overall picture were different in each room.

In room 1, students embodied an intrinsic patch perspective. This arrangement created "constrained but open-ended" (Goodwin, p. 32, 2018) possibilities for student and instructor interactions during the activity. From this intrinsic perspective students could not see the composite image made by all of the patches, which meant that it was not necessary from the students' perspective to show their patch color uniformly. It is important to note that these students held their colored squares for the entirety of the activity and all presented them in some manner after they responded to the Observer, which was easily understood as part of their role as patches. As shown in Figure 2, students in room 1 did not present their colored squares uniformly. There are three students visible in Figure 2 that all presented their colored squares in different spatial orientations, as if to different audiences. The first student (1) held her patch in her left hand, presenting the orange square on the other side to someone standing on her right. From the camera's perspective it looks like she is presenting a red square, however, her rightward gaze indicates where or to whom she was presenting her colored square. The boy to her right (2) held his colored square above his head, presenting an orange square towards the instructor or Observer and the girl to his right (3) held her orange square facing outwards from her chest. While all three of these students held their colored squares orthogonal to the ground, some students even held their squares above their heads, parallel to the ground, visible from a bird's eye view. This meant that it was hard for participants both intrinsically and extrinsically to easily see the color of each patch or the resultant image. This also meant that it made it more difficult for students as patches to look to their neighbors for clues as to how to respond to a prompt, something that patches cannot do either. Similar to the conflation of person and role with respect to the

Observer as described in the previous section, in room 1 the person and role of patch was also tightly coupled due to the students' physical location inside the grid standing in as their patches for the entire activity.



Figure 2. Students stood in different places on the grid, which afforded different visibilities and perspectives.

In room 2, students stood outside of the grid, which meant that they had to take an intrinsic perspective (to think like a patch) while standing in an extrinsic position in which they could see the entire image (something that a patch could not “see”). From this position students in room 2 did not hold their squares the entire time, but they still intuitively understood their relationship to the colored squares, manipulating them when asked by the observer and then setting them down in the grid. This meant that the composite image created by the students in room 2 was visible (Figure 2), which made it possible for the group to evaluate their image. Had all of the students in room 2 (Figure 2) responded to the Observer correctly the resultant image would have been two stripes. In Figure 2 it is clear that one of the students in the orange stripe had her colored square blue instead of orange. This visible difference became a point of discussion for the class, prompting Neil to talk through interpreting the code with the student and later led Kiara to ask her what had been confusing. It is also important to note that visual designs were visible in this small grid to the extent that patterns (e.g. stripes) were easily accessible to the group. The group began to manipulate and see the visual effects transform the gridded canvas in a way that emphasized that the image consisted of individual programmable agents (patches).

Not only was the composite image visible, but student manipulation of colored squares was visible as well. It was clear when students were changing their patch color and when they were done (Figure 3). This meant that instructors could tell which patches thought they were being talked to and which were not, leading them to reiterate what “ask patches,” meant (“ask ALL patches”). Similar to the way Kiara positioned the Observer as a messenger, patches in room 2 consisted of a system of students separate from their colored squares. This consistent social structure made processes and creations visible, which is not afforded in the computational environment of NetLogo. In both rooms, the relational metaphor of the Observer-patch relationship was further embodied by where students stood on the grid, continuing the development of differing substrates which influenced how students interacted with the Observer’s code and the squares they manipulated.



Figure 3. In room 2, it was clear which students responded to a prompt and how they responded.

### Order of translation

The differing substrates in each room were also influenced by the order in which facilitators translated between NetLogo code and colloquial English when presenting code for students to interpret as patches. This led to the development of different frames for enacting code. Tracy used an *encoding* frame: she began with colloquial English and then translated her commands into NetLogo code. Kiara used a *decoding* frame: she began by asking students to interpret NetLogo code and once it was performed she translated it into colloquial English.

In room 1, Tracy began by describing the first command to the patches as a desired process in plain English words and praised students for quickly understanding the code, although she had not yet discussed the

specific syntax needed for the computational patches in the computer to respond:

So as the Observer, I might be thinking that I want to ask you all if you are black set it to red. So that might be something that I want to say and so since you all are, *((students manipulated their colored squares))* since you guys are all so smart and know both English and how we speak when we code, you guys already knew that if you were black you went ahead and set it to red. Did anybody not set it to red? No? All right, so the way we might ask this in coding language is, as the Observer, so I'm the Observer right now, I would say, "ask patches [if pcolor = black [set pcolor red]]"

Again, Tracy began by unifying herself as the Observer and commander, "I might be thinking I want to ask you." It was Tracy's agency as a human Observer that dictated what the students as patches had to do, the idea came from her. The students quickly responded by changing their colored squares to red, although nobody questioned whether they had started with black squares. Students were initially handed their squares with black on one side; however, students explored flipping through their colored squares on their own so by the time Tracy called out this command not all students had their colored squares in the same orientation—another challenge due to their location and position in the grid. It is also unclear how Tracy quickly asserted that all students correctly changed their colored squares to red, but her assertion, "since you guys are all so smart and know both English and how we speak when we code, you guys already knew that if you were black you went ahead and set it to red," was a pedagogical move in which Tracy gave students early, low-entry-level success evident in their quick manipulation of their colored squares. Speaking in an English translation of the NetLogo code at first and asserting students interpretations as a correct positioned students as successful from the beginning before they started to encode the NetLogo translation of the same command.

In room 2, Kiara and Neil gave the same command to their students' patches by first saying the specific NetLogo code and then decoding its meaning, framing coding as a process of interpretation:

Kiara: So it tells me to ask the patches, so here goes, "[if pcolor = black [set pcolor red]]"

Neil: What do you think pcolor means? Patch color. So if patch color equals black, does anyone body think they might have a patch color that is black? Right?

Kiara: What would set pcolor red mean? What do you think you'd have to do?

Student: Turn it red.

Kiara: Turn it to the red.

Neil: All right.

*((Students manipulated their colored squares, one row changing their black squares to red.))*

Kiara: So in everyday life we don't go around saying all the, all the people holding black just turn pcolor, turn the equal black now I want you to set pcolor red. Is that how we speak in everyday language?

Students: No

Kiara: No! That is our computer language, but if we translate that, what I actually said was "ask all patches," which is just this side right *((points to the row of students who changed their colored squares))*? All patches is just this side right?

Student: No

Kiara: NO! It's who?

Student: All of these

Kiara: All of you! But I was only talking to which color first?

Students: Black

Kiara: Black. So it says ask all patches, which I asked everybody first, and then I said, "if your color is black," so you immediately looked down to see if you had black and it, I said, "set it to red," and they changed their cards to red.

Kiara began again by positioning herself in the Observer role as a messenger to the patches, "so it tells me to ask the patches." This cued a syntax detail in the NetLogo code; "ask patches," which syntactically positioned what followed as a request or command spoken to patches to follow. Kiara and Neil continued by talking students through interpreting the new vocabulary and syntax. Right away, Neil chimed in asking students to

interpret new vocabulary, asking them what pcolor meant and whose pcolor was black. Kiara continued by asking students what setting pcolor red might mean. Once the students had interpreted the code they then proceeded to perform it: all of the students whose colored squares were black turned them to be red. Breaking down the potentially intimidating code into smaller, digestible interpretative questions (e.g. What does pcolor mean? What does set pcolor red mean?) facilitated students' early success interpreting their first line of code.

Kiara emphasized that although NetLogo code is not how people normally talk, it is comprehensible from an embodied patch perspective, emphasizing the ego syntonicity (Papert, 1980) between students and patches. Through this translation process she made certain aspects of the syntax salient to the students, emphasizing that although not all students changed their pcolors, they were all spoken to because "ask patches" meant "ask all patches." Noting that the first step was to attend to noticing if one's patch color was black—"so you immediately looked down to see if you had black"—was an important embodied interpretation of the code, positioning *all* individual computational agents as needing to interpret the coded request. In room 2, not only was the instructor playing the role of the Observer treated as a messenger, but the students themselves were a distributed part of a patch-person-square-system as messengers to their colored squares. This decoding frame positioned students as interpretive agents translating NetLogo syntax into actionable English that they understood as humans in contrast to the encoding frame that developed in room 1 in which ideas were first expressed in English by the agentic instructor/Observer and then translated back into NetLogo.

### Differences in students' subsequent activity

The different structural pedagogical moves that instructors made in each room created differing substrates in which two predominant student coding dispositions developed. Goodwin (2018) writes that substrates structure subsequent activity and thus students' participation as the Observer provides evidence that different substrates sedimented. Although students in both rooms engaged playfully as they took on the role of the Observer, they utilized the different frames that developed in each room. In room 1, students engaged in an *encoding frame* by first sharing complex ideas for the composite image that eventually over extended even the instructors' computational understandings, stretching the group's encoding abilities. In contrast, the students in room 2 engaged in a *decoding frame* in which they proposed much simpler code to the group, but communicated their code directly using discourse that more closely resembled NetLogo syntax. We argue that room 1 enacted an encoding frame for understanding code in which one started from open-ended possibilities that could be articulated in English and room 2 enacted an interpretive frame for understanding code in which interpretation of a new language was foregrounded. Not only were different dispositions for how to code developing in these two rooms, but the students that chose to participate as Observers differed in both rooms. In room 1, the most vocal students, three boys with coding experience, volunteered to play the Observer, while in room 2 boys and girls all of whom had no previous coding experience participated as Observers. The different substrates that developed in each room influenced the subsequent activity as evidence in students' enactments of the Observer.

### Student observers generated different prompts in each room

When students were given the opportunity to perform the part of the Observer in this physical ABM environment, students in each room consistently enacted the different coding frames that had developed in each room. Although both instructors asked students to try out their own ideas as the Observer, students in each room took this up either as *encoders*, sharing their ideas first in English, or as *decoders*, sharing their ideas first in NetLogo. In room 1, Tracy emphasized an encoding frame, telling each student Observer to first tell the group what he wanted to happen and then make the same command in NetLogo: "So Kevin, you tell us what you want to happen and then we're gonna help you translate it into a language that the patches understand." When student Observers shared their ideas in English, before attempting to encode them in NetLogo, student patches immediately started responding by manipulating their colored squares. When Charlie began as the Observer, he shared with the class, "Ok, so first I want you to pick a random number from 0 to 5," and students immediately began flipping through their colored squares. The social relationship between the Observer and patches was so strong that as soon as the Observer started uttering commands, the patches immediately responded. Tracy, however, quickly asked them to stop, "Ok, hold up patches, hold up. He was just telling in English first." Tracy's request that they wait for the official code ran counter to the students' understanding of their role, and emphasized that patches could only understand the specifics of coding languages. Tracy also wrote the NetLogo code on the whiteboard as students developed the code out loud, providing another representation of the code and emphasizing the official syntax as what patches could respond to. This sequence of action was dependent on the substrate that began to sediment throughout the activity as described previously.

Similar to how Tracy positioned the student Observers in room 1, Kiara also encouraged the student Observers to try out their own ideas with the group: "Are there any other ideas you want to try? Anybody want

to be the Observer? ((One student raised his hand.)) Ok, you want to be the Observer? Ok, you go and I'll take your place. And you tell us what you want us to do." Kiara asked students to share their ideas, not specifying that they had to be in NetLogo or in English as Tracy had. The student Observers in room 2, however, all talked in a manner that resembled NetLogo code. Their syntax was not perfect, but it was also clear that they were not talking in colloquial English like the students in room 1. In contrast to student Observers in room 1, the students in room 2 spent about half as much time in the Observer box, quickly reciting their code to their peers. The barrier for entry seemed much lower in room 2 and although the ideas the student Observers tried were simpler than in room 1, students' ability to speak in this new coded language and decode it as patches was much more fluent. As substrates developed and sedimented in each room, students' understandings of how to code in this ABM remained consistent with the encoding or decoding frame they established, reinforcing these dispositions.

## Discussion and conclusion

Facilitation of these Stadium Card Activities appears to afford a wide range of choices. Even with tightly scripted and explicitly documented materials, we found that our teacher partners could make divergent and separately coherent patterns of facilitation choices, that produced (a) distinct opportunities for student participants to experience intrinsic and extrinsic perspectives on their shared simulations, and (b) distinct images of the nature of agent-based commands. Though our analysis is preliminary, these different facilitation choices also appeared to be consequential for the kinds of creative agent-based programming ideas that students generated in the moment. In both cases, the activities succeeded in fostering creative expression, but there was a distinct flavor to each of the two rooms, resonant with the choices the facilitators made. The social and embodied performances of the infrastructure of an ABM environment was an accessible way for learners to grapple with important computational ideas with a lively tone that invited expressive play. In our next design iteration we hope to explore students' participation in a similar activity that allows students to leverage both intrinsic and extrinsic perspectives and explicitly acknowledges these the two frames for coding (encoding and decoding) that we saw develop in this first design iteration.

In addition, creating visual effects on an embodied patch matrix at a small scale provided learners with a "mid-level" (Levy & Wilensky, 2008) conception of expressive designs. The Stadium Cards provided an activity-type for thinking with patches, which addresses core ideas in that area. These activities can be seen as providing experiences that fill an analogous role to that played by basic work in turtle graphics, for turtles (movable agents). To fulfill an expressive function, patches must be thought of in groups, and recognizable visual forms only become coherent with an enormous number of patches (think of the unit of the megapixel), but a mid-level scale of patches is accessible to learners and gives them an entry point to think about dynamic, computational images as composed of a grid of individually programmable patches/pixels.

## Endnote

- (1) In these transcripts, whenever anybody read NetLogo code aloud we transcribed their verbalization as the written code itself. This was done to emphasize the syntactical differences between NetLogo code and colloquial English.

## References

- Bontá, P., Papert, A., & Silverman, B. (2010). Turtle, art, turtleart. *Proceedings of Constructionism 2010*.
- Fine, G. A. (1979). Small Groups and Culture Creation: The Idioculture of Little League Baseball Teams. *American Sociological Review*, 44(5), 733. <https://doi.org/10.2307/2094525>
- Goodwin, C. (2017). *Co-operative action*. Cambridge University Press.
- Levy, S. T., & Wilensky, U. (2008). Inventing a "mid level" to make ends meet: Reasoning between the levels of complexity. *Cognition and Instruction*, 26(1), 1-47.
- Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas*. Basic Books, Inc..
- Resnick, M., Maloney, J., Monroy-Hernández, A., Rusk, N., Eastmond, E., Brennan, K., ... & Kafai, Y. (2009). Scratch: programming for all. *Communications of the ACM*, 52(11), 60-67.
- Resnick, M., & Wilensky, U. (1998). Diving into complexity: Developing probabilistic decentralized thinking through role-playing activities. *The Journal of the Learning Sciences*, 7(2), 153-172..
- Wilensky, U. 1999. NetLogo. <http://ccl.northwestern.edu/netlogo/>. Center for Connected Learning and Computer-Based Modeling, Northwestern University. Evanston, IL.

## Acknowledgments

This material is based upon work supported by the National Science Foundation under No. 1742257.

# Understanding the Effect of Group Variance on Learning

Sinan Yavuz, Michael Forkner, Dana Gnesdilow, Catherine L. Dornfeld Tissenbaum, Jee-Seon Kim, Sadhana Puntambekar

syavuz@wisc.edu, forkner@wisc.edu, gnesdilow@wisc.edu, catdoti@illinois.edu, jee-seonkim@wisc.edu, puntambekar@education.wisc.edu

Department of Educational Psychology, University of Wisconsin-Madison

**Abstract:** Given that group composition is a key factor that affects learning in CSCL environments, it is important to study how students in groups with homogeneous or heterogeneous levels of prior knowledge collaborate. This study investigated the potential differences in students' learning outcomes from participating in a 13-week design-based unit. We used the pre- and post-test data from 361 eighth grade students (102 groups) and performed a hierarchical linear model analysis to examine how the convergence or divergence in the students' level of prior knowledge affected students' learning outcomes. We found that students in homogenous groups with similar levels of initial prior knowledge scored significantly higher on their post-test when their pre-test was used as a covariate, than students in heterogeneous groups. Implications of these findings are discussed along with directions for future research.

## Introduction

Group composition is a key factor affecting learning in CSCL (Puntambekar & Young, 2003). Often, group members come with different levels of prior knowledge, bringing convergence or divergence of ideas (Weinberger, Stegmann, Fischer, 2007). It is therefore important to study how student groups with homogeneous or heterogeneous levels of prior knowledge collaborate. Vygotsky's (1978) Zone of Proximal Development (ZPD) framework assumes that there is a "more capable other" who can scaffold a learner to accomplish more with assistance than alone. The ZPD therefore implies that for peers to support and scaffold learning in group interactions, there has to be an inherent asymmetry in the group's knowledge. Stahl (2004) also suggested that divergent ideas are an essential mechanism for the exploration of ideas and negotiation of knowledge during group collaborations.

However, in classrooms, groups are often composed of students with more homogeneous than heterogeneous levels of prior knowledge. It may also be the case that heterogeneous groups mimic homogeneous groups, as the more capable other does not actually provide the proper support to the other members. Because of the homogeneous nature of these groups, multiple students may collectively scaffold each other, rather than only the more capable peer providing the scaffolding. For example, Fernández, Wegerif, Mercer and Rojas-Drummond (2001) found that students' dialogue during collaboration in symmetrical (i.e., homogeneous) groups provided enough support to help students solve problems, and thus argued for a reconceptualization of the relationship described by Vygotsky's ZPD. Further, other researchers have claimed that peers may not intentionally try to scaffold each other; but, by working together, peers can solve a problem or complete a task that they could not accomplish when working alone (Wells, 1999; Zuckerman, 2003).

Prior research that examined homogeneous and heterogeneous group composition identified learning benefits when students were placed in heterogeneous groups (Csanadi, Kollar, & Fischer, 2016; Gijlers & De Jong, 2005; Patchan, Hawk, Stevens, & Schunn, 2013; Wiedmann, Leach, Rummel, & Wiley, 2012; Zhao et al., 2018). These benefits may be due to the processes that occur as students work together during collaborative knowledge-building activities, as peers with greater knowledge bring important issues and resources, while peers with less knowledge play an important role by raising questions and asking for clarifications, which the peers with greater knowledge may address (Scardamalia & Bereiter, 1994).

Yet other researchers have found that, in some cases, homogeneous group composition is preferable. Webb, Nemer, and Zuniga (2002) found that high-ability students performed better in homogeneous groups, whereas low-ability students performed better when they had a more capable other in their heterogeneous group. Along these lines, Lou et al. (1996) found that high-ability students benefited equally from both homogeneous and heterogeneous groups, while medium-ability students benefited most from homogeneous groups, and low-ability students benefited most from heterogeneous groups. Other research has identified that same-ability dyads were better at metacognitively regulating their collaborative process to reach their goals than their heterogeneous partners (Zillmer & Kuhn, 2018).

To further understand the relationship between groups with similar or different levels of prior knowledge, our study examined 102 groups of middle-school students learning science over 13 weeks. Students worked with the same group as they engaged in CSCL tasks each school day, which gave us the unique opportunity to examine how group composition, based on levels of prior knowledge, affected students' conceptual learning outcomes, using a larger sample over a longer time frame than many prior studies.

## Methods

### Participants and instructional context

This study took place during the 2016-2017 academic year, with seven science teachers and their 515 eighth-grade students (229 female and 286 male). All students and teachers were from one of three middle schools in the same urban school district in the U.S. Midwest. This district served about 2,066 middle school students, with about 53% of them identified as being economically disadvantaged. Students in all classes participated in a design-based unit called "Make Your Own Compost!". The curriculum challenged students to create a compost that would break down quickly and contain nutrients while minimizing landfill waste and other negative effects of conventional fertilizers on the environment. Students collaborated in the same group of three to six students (mean group size is 3.54) to learn about ecosystems, energy transformations, matter cycling, and human impacts over the entire 13 weeks of the unit. Students participated in a variety of science activities, such as experiments, and worked collaboratively using computers to conduct research using an online digital text. They also ran multiple compost simulations to help them to build the necessary knowledge to solve the challenge over the course of the unit. All activities in the unit were designed to help students to solve the challenge and write a final report to their principal to propose their design of a composting program for the school. For this study, we examined the pre- and post-tests from 150 groups of students. However, due to missing data and varying group sizes, we only included groups for which we had both pre- and post-test information from at least three students in a group. The results of this study are based on data from 102 groups (361 students), who were assigned by the teacher. Each student took the pre-test prior to being introduced to unit content and activities. Students took the post-test after finishing the unit. Groups were categorized as homogeneous, medium, and heterogeneous based on their pre-test score variances.

### Data sources and analysis

#### Pre- and post-test measures

The "Make Your Own Compost!" unit focused on helping students to build science understanding about ecosystems and humans' impact on them. The test was designed by the research team and consisted of 24 questions that assessed students' understanding of concepts and relationships related to biotic and abiotic factors in ecosystems; organisms' roles and relationships in ecosystems; the flow of energy and cycling of matter in ecosystems; and human impacts on ecosystems. Four of the 24 questions were open-ended items, and 20 questions were in a multiple choice (MC) format. Three of the 24 questions had multiple parts. Overall test reliability was calculated using Cronbach's alpha. We found the test to be reliable with alpha values of .863 for the pre-test and .874 for the post-test. After analysis of the open-ended items, a conflict was determined between the pre- and post-test scores due to incomplete responses on the post-test. Therefore, we only included MC items in our analyses. The maximum score students could earn on the MC items was 23.

#### Hierarchical Linear Models

In this study, students were nested in groups, and we could not assume that students' learning gains were independent, as the intervention was applied in a group setting. Thus, students' learning gains within the cluster were expected to be correlated, and the dependency between students needed to be considered (Kim, Anderson, & Keller, 2013). Therefore, using classical statistical methods, such as linear regression or ANOVA on the student-level data, while ignoring the group clustering effect, would lead to inaccurate results and interpretation. By using hierarchical linear models (HLM), we could represent each level by its own submodel (Raudenbush & Bryk, 2002) to consider the multilevel nature of the data. We performed all the analyses using R software (R Core Team, 2017) and used the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015) for HLM and *ggplot2* for producing graphs (Wickham, 2009). We considered student level as level-1 and group level as level-2.

## Results



Since we were interested in investigating group variability, we only analyzed groups with three to six students who completed both the pre- and post-test. After eliminating groups that did not fit our criteria, we were left with 361 students who were nested in 102 groups. As mentioned before, we only analyzed the students' total score for the MC questions. Descriptive statistics of students' pre- and post-test item scores are shown in Table 1.

Table 1: Level-1 Variables (N = 361 Students)

| Variable         | N   | Mean  | Std Dev | Median | Max. Poss. |
|------------------|-----|-------|---------|--------|------------|
| Pre-Test Scores  | 361 | 15.05 | 4.43    | 15.5   | 23         |
| Post-Test Scores | 361 | 19.46 | 3.82    | 21     | 23         |

We divided the student groups into three group types based on their pre-test score variance, which ranged from 0.25 to 72.58. Thirty-three percent of the groups with the lowest variance were classified as homogenous; the middle 34% were classified as medium; and 33% of the groups with the highest variance were classified as heterogenous. The variance of within-group prior knowledge distribution is shown in Figure 1. The descriptive statistics of the groups are shown in Table 2.

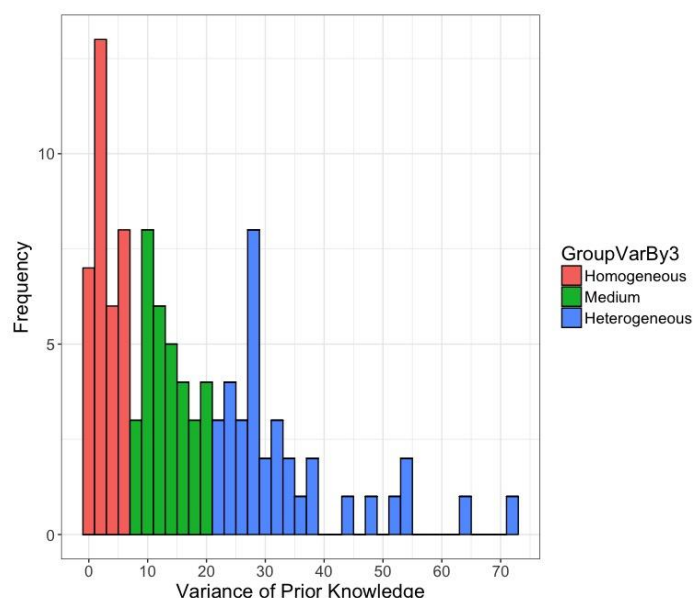


Figure 1. The histogram of within group prior knowledge variance.

Table 2: Level-2 Variables (N = 102 Groups)

| Variable         | Levels        | N  | Mean  | Std Dev | Median |
|------------------|---------------|----|-------|---------|--------|
| Pre-Test Scores  | Homogeneous   | 34 | 15.21 | 3.31    | 15.19  |
|                  | Medium        | 33 | 15.82 | 2.70    | 16.33  |
|                  | Heterogeneous | 35 | 14.07 | 2.08    | 14.42  |
| Post-Test Scores | Homogeneous   | 34 | 19.97 | 2.53    | 20.35  |
|                  | Medium        | 33 | 19.87 | 2.05    | 19.94  |
|                  | Heterogeneous | 35 | 18.60 | 2.02    | 18.12  |

Figure 2a shows student-level pre- and post-test scores based on their group variability. From this graph, we can see that students whose groups we designated as homogenous showed a steeper line from pre- to post-test, meaning they had the highest learning gains of all three group types. Figure 2b shows group level means for homogenous, medium, and heterogeneous groups. Again, we see that the homogenous level groups have comparatively steeper lines, showing they had higher average learning gains than the other two groups. These differences were found to be statistically significant, which we describe below.

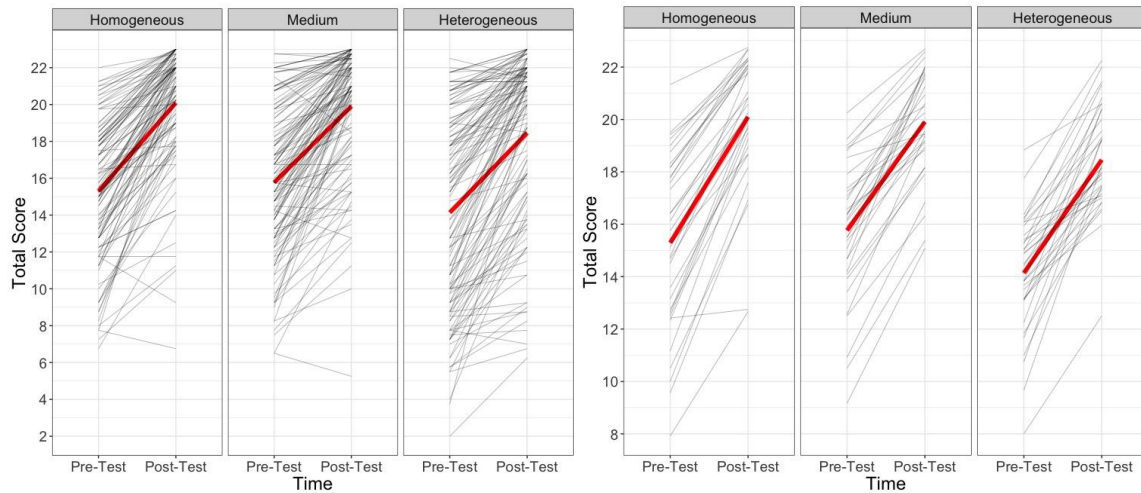


Figure 2 a & b. Student- (left) and group- (right) level pre- and post-test scores.

### HLM analyses

We ran HLM analyses to investigate whether being in a homogenous, medium, or heterogenous prior knowledge group at the start of the unit affected students' learning gains. We used students' pre-test scores as a covariate and post-test scores as an outcome variable. Then we added group variability as an independent variable to the model so that we could investigate whether being in a homogenous, medium, or heterogenous prior knowledge group affected students' learning gains.

Table 3: HLM Model coefficients, standard errors

| Fixed Effect                 | Coefficient (SE) | t (df)         | p      |
|------------------------------|------------------|----------------|--------|
| (Intercept) - $\gamma_{00}$  | 10.28 (0.54)     | 19.17 (309)    | 0.0000 |
| PreScore - $\gamma_{10}$     | 0.64 (0.03)      | 20.76 (355.45) | 0.0000 |
| Medium - $\gamma_{20}$       | -0.58 (0.36)     | -1.58 (100.48) | 0.1163 |
| Heterogenous - $\gamma_{30}$ | -0.72 (0.36)     | -1.99 (99.96)  | 0.0495 |
| Variance Components          | Estimate         |                |        |
| Residual - $\sigma^2$        | 6.01             |                |        |
| (Intercept) - $\tau_{02}$    | 0.49             |                |        |

In this model, we added groups' homogeneity level as an independent variable, using dummy coding. The model can be written as follows:

|              |   |
|--------------|---|
| Level-1:     | $PostScore_{ij} = \beta_0 + \beta_1 PreScore_{ij} + \beta_2 Medium + \beta_3 Heterogeneous + R_{ij}$                  |
| Level-2:     | $\beta_0 = \gamma_{00} + U_{0j}$<br>$\beta_1 = \gamma_{10}$<br>$\beta_2 = \gamma_{20}$<br>$\beta_3 = \gamma_{30}$     |
| Mixed Model: | $PostScore_{ij} = \gamma_{00} + \gamma_{10} PreScore_{ij} + \beta_2 Medium + \beta_3 Heterogeneous + R_{ij} + U_{0j}$ |

Based on the results in Table 3, we rewrote the model equation as:

$$PostScore_{ij} = 10.28 + 0.64 PreScore_{ij} - 0.57 Medium - 0.72 Heterogeneous$$

After controlling for pre-test scores, the average adjusted post-test score for students in homogeneous groups was 10.28. Additionally, students' average adjusted post-test score in the medium and heterogeneous groups were 9.70 and 9.56 respectively. We found that students in homogeneous group performed significantly better on the post-test than the students in the heterogeneous group ( $t(99.96) = -1.99$ ,  $p = 0.0495$ ), but not the medium group ( $t(100.48) = -1.58$ ,  $p = 0.1163$ ). Analysis shows intraclass correlation as .08, which means that group membership explains 8% of the variance in the post-test scores.

Because we were interested in knowing whether students with lower levels versus higher levels of prior knowledge benefitted similarly from being in a homogenous group, we ran further analyses to examine the

difference within the homogeneous group. To do this, we examined the range of all students' pretest scores and then divided the homogenous groups into thirds to make three pre-test score categories: low ( $n = 11$  groups), medium ( $n = 11$  groups), and high ( $n = 12$  groups). The adjusted average post-test scores of the students in these three categories were 10.11, 10.30, and 10.07 respectively. After running linear regression analyses, we found no significant differences between the three categories of students within the homogeneous group, when controlling for pre-test scores.

## Discussion

We conducted a quantitative study to explore students' learning based on whether they differed in their levels of prior knowledge, in order to understand whether group composition based on students' prior knowledge was a factor affecting collaboration (Puntambekar & Young, 2003; Weinberger, Stegmann, & Fischer, 2007). Based on Vygotsky's conceptualization of the ZPD, many educators have historically believed that students should be placed in heterogeneous groups so that more capable students are able to support their less capable peers. More capable students in heterogeneous groups can provide knowledge and resources and address questions that less capable peers raise (Scardamalia & Bereiter, 1994). However, we found that students in similar-prior knowledge groups (i.e., homogeneous) had significantly higher learning gains than students in mixed-prior knowledge groups (i.e., heterogeneous) after working together during a design-based unit over 13 weeks. These findings contrast with those of other researchers who have found that students in heterogeneous groups showed greater learning outcomes than students in homogeneous groups (e.g., Csanadi et al., 2016; Gijilers & De Jong, 2005; Patchan et al., 2013; Zhao et al., 2018).

Yet, our findings also indicated benefits for students working in homogenous groups. We found that students in homogenous groups with low, medium, or high prior knowledge benefited equally from collaborating with similar level peers. These results differ from both Webb, Nemer, and Zuniga (2002) and Lou et al. (1996), who found that low prior knowledge students performed better in heterogeneous groups, while high prior knowledge students performed better, or equally as well, in homogenous groups. However, we only analyzed students' pre- and post-test scores on a content test in our study. In the absence of analysis of the discourse among group members, we cannot explain why these results occurred. However, one possible explanation, based on Zillmer and Kuhn's (2018) findings, is that students with similar prior knowledge levels may be more capable of providing metacognitive support to their peers, switching roles dynamically as needed. Zillmer and Kuhn (2018) also pointed out that students with similar levels of prior knowledge may have been better at metacognitively scaffolding one another because they worked together for a greater amount of time, which helped them establish greater intersubjectivity. In our study, students worked in groups for 13 weeks, which is a relatively long period of time. It could be the case that students in homogenous groups were better able to establish intersubjectivity earlier and maintain it as they collaborated throughout the unit.

Another possible explanation is that the more capable peers often do not intentionally scaffold other group members (Wells, 1999; Zuckerman, 2003). Analysis of students' discourse will help us understand the extent to which students in homogeneous and heterogeneous groups provided scaffolding, or not. Our findings lend support to Fernández and colleagues' (2001) argument that Vygotsky's ZPD framework should be reconceptualized to capture the kinds of scaffolding that occur in symmetrical groups. For example, different types of dialogue can provide support for symmetrical groups, which may help them to pool the groups' intellectual resources and establish a common context to collaboratively solve a problem that they would not be able to solve individually. In symmetrical groups, different group members might provide scaffolding at different times, in a continuously unfolding, reciprocal, and ever-changing process, as participants discursively contribute to the collaboration.

## Future research

We plan on analyzing students' discourse from the video and audio data collected during the 13-week unit. This will help us to understand the types of contributions and support that students provided each other within their groups and try to determine why homogenous groups had higher learning gains than the heterogeneous groups in our study. Further, Zillmer and Kuhn (2018) suggested that the length of time students spend collaborating impacts the quality of their interactions. While we collected data over a longer time period than many previous studies, future research could also examine how the length of collaboration, a few sessions to weeks or months of collaboration, affects students' learning outcomes based on whether they are in a homogenous or heterogeneous prior knowledge group. Finally, future research could also examine how student's learning gains differ within the groups at the end of the unit. This would help us to identify the extent to which students starting out with different levels of prior knowledge are benefitting from participating in each type of group and

whether students' understanding becomes more convergent over time. Information from each of these lines of research could provide practical guidance to teachers as they are strategically forming groups in the classroom.

## References

- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48.
- Csanadi, A., Kollar, I., & Fischer, F. (2016). Scientific Reasoning and Problem Solving in a Practical Domain: Are Two Heads Better Than One?. *Singapore: International Society of the Learning Sciences*.
- Fernández, M., Wegerif, R., Mercer, N., & Rojas-Drummond, S. (2001). Re-conceptualizing "scaffolding" and the zone of proximal development in the context of symmetrical collaborative learning. *The Journal of Classroom Interaction*, 40-54.
- Gijlers, H., & De Jong, T. (2005). The relation between prior knowledge and students' collaborative discovery learning processes. *Journal of Research in Science Teaching*, 42(3), 264-282.
- Kim, J. S., Anderson, C. J., & Keller, B. (2013). Multilevel analysis of assessment data. Rutkowski, L. von Davier, M., & Rutkowski, D. (Eds.), *Handbook of international large-scale assessment: Background, technical issues, and methods of data analysis*, 389-425.
- Long, J. D. (2012). *Longitudinal data analysis for the behavioral sciences using R*. London, England: Sage Publication.
- Lou, Y., Abrami, P. C., Spence, J. C., Poulsen, C., Chambers, B., & d'Apollonia, S. (1996). Within-class grouping: A meta-analysis. *Review of Educational Research*, 66(4), 423-458.
- Patchan, M. M., Hawk, B., Stevens, C. A., & Schunn, C. D. (2013). The effects of skill diversity on commenting and revisions. *Instructional Science*, 41(2), 381-405.
- Puntambekar, S. & Young, M. F. (2003). Moving toward a theory of CSCL. In U. Hoppe, B. Wasson, & S. Ludvigson, (Eds.) *Computer Supported Collaborative Learning 2003: Designing for change in networked learning*. pp. 503-512. IOS Press: Amsterdam. [Presented in the closing plenary session at CSCL 2003].
- R Core Team (2017). R: A language and environment for statistical computing [Computer software]. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). (2nd ed.). Thousand Oaks: Sage Publications. Sage Publication.
- Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context*. Oxford, England: Oxford University Press.
- Scardamalia, M., & Bereiter, C. (1994). Computer support for knowledge-building communities. *The journal of the learning sciences*, 3(3), 265-283.
- Stahl, G. (2004). Building collaborative knowing. In J.W. Strijbos, P.A. Kirschner, & R.L. Martens (Eds.), *What we know about CSCL* (pp. 53-85). Springer, Dordrecht.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher mental process*.
- Webb, N. M., Nemer, K. M., & Zuniga, S. (2002). Short circuits or superconductors? Effects of group composition on high-achieving students' science assessment performance. *American Educational Research Journal*, 39(4), 943-989.
- Weinberger, A., Stegmann, K., & Fischer, F. (2007). Knowledge convergence in collaborative learning: Concepts and assessment. *Learning and Instruction*, 17(4), 416-426.
- Wells, G. (1999). *Dialogic inquiry: Towards a socio-cultural practice and theory of education*. Cambridge University Press.
- Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis* Springer-Verlag New York.
- Wiedmann, M., Leach, R. C., Rummel, N., & Wiley, J. (2012). Does group composition affect learning by invention?. *Instructional Science*, 40(4), 711-730
- Zillmer, N., & Kuhn, D. (2018). Do similar-ability peers regulate one another in a collaborative discourse activity?. *Cognitive Development*, 45, 68-76.
- Zhao, J., Lin, L., Sun, J., Zheng, X., & Yin, J. (2018). Students' engagement in a science classroom: Does knowledge diversity matter?. *The Journal of Educational Research*, 1-8.
- Zuckerman, G. (2003). The developmental path toward reflection. In A. Kouzulin, B. Gindis, V.S. Ageyev, & S.M. Miller (Eds.), *Vygotsky's educational theory in cultural context*, 177. Cambridge.

## Acknowledgments

The research reported in this study is supported by a NSF-DRL grant (#1418044) to the last author. We thank all of the teachers and students who participated in this study.

# Arguing about Synthetic Biology in 140 Characters or Less: Affordances of Microblogging for High School Students Discussions of Socio-Scientific Issues

Mia S. Shaw, Justice T. Walker, and Yasmin B. Kafai  
mshaw12@gse.upenn.edu, justicew@gse.upenn.edu, kafai@upenn.edu  
University of Pennsylvania

**Abstract:** As synthetic biology becomes more prevalent, there is a call to critically and publicly evaluate the benefits and risks of such technologies. Microblogs like Twitter have the potential for engaging citizens in civic discourse about such issues but also pose challenges due to partisan behaviors and length limitations. Given the saliency of social media in young people's lives, we examined in an exploratory study whether integrating a 140-character limit within an audience response system supported high school students' argumentation about socioscientific issues during a synthetic biology workshop. After analyzing students' anonymous online responses, worksheets, and observations, we found that students generated diverse evaluative claims as well as used argument-based rationales and weighed counter arguments when assessing their peers' claims. In our discussion, we address the affordances and constraints of microblogging in supporting students' argumentation of socioscientific issues about synthetic biology.

## Introduction

Biotechnologies such as synthetic biology have increasingly moved out of research labs and are now more prevalent in our everyday lives and in industries that impact medical, environmental and textile industries. Understood as an engineering discipline, synthetic biology builds complex biological systems from standard, interchangeable biological parts (Cheng & Lu, 2012). In other words, synthetic biology involves the manipulation and design of organisms and/or their outputs. Given the myriad impacts synthetic biology has on society and the environment, a variety of stakeholders including experts, policymakers, and citizens have made evaluating the potential benefits and risks of these applications more salient (Gutmann, 2011). Within a similar vein, there is a call for science education to provide opportunities for students to engage in argumentation around claims developing scientific fields assert in order to simulate the discursive practices observed within scientific communities and in the world at-large (Driver, Newton, & Osborne, 2000). As synthetic biology technologies emerge in K-12 education, students will need to be able to engage in argumentation as future decision-makers and citizens. To this end, they will need to also consider the potential benefits and risks of these emergent technologies and their impact on humans, nonhumans, and the environment.

Microblogs—online platforms where users share and respond to brief 140-200-character messages—and other social media have influenced public awareness about science-related matters and their risks (Auer, Zhang, & Lee, 2014). Even though scientists have criticized using microblogs (e.g., Twitter) for communicating with the public (Shäfer, 2012), these platforms afford discourse wherein individuals witness and participate in diverse conversations they otherwise may not have before (Yardi & boyd, 2010). As Twitter use among young people becomes more and more prominent (Greenhow & Gleason, 2012), research has pointed to the potential for microblogs to serve as platforms for socioscientific (SSI) argumentation skill development among students (Greenhow, Gibbons, & Menzer, 2015). A distinguishing feature of microblogs is that as messages (e.g. “tweets” on Twitter) from different individuals aggregate, their collective whole can form a coherent text where narratives emerge (Murphy, 2013). Considering that wording choices on microblogs can shape how readers process and share information, there exists an opportunity to explore how microblogs promote (or inhibit) science-based discourse (Brossard, 2013). Furthermore, given the character limitation of microblogs, it is important to consider whether or not argumentation of complex SSIs can take place on such a constrained setting.

In this exploratory study, we examine the various ways high school students in a synthetic biology-based workshop discuss SSIs surrounding the field's applications within a microblogging context. In a weeklong summer science workshop, students genetically transformed yeast cells to produce beta-carotene and used this new genetically-modified yeast to produce vitamin-enriched cakes (hereafter, referred to as Biocakes). Following, we facilitated a class-wide discussion about three different synthetic biology applications in the food, medicine, and textiles industries using Mentimeter, a free and existing commercial audience response system

(ARS) that polled and displayed students' anonymous responses. Due to the 140-character limitation on Mentimeter responses, students critically examined these synthetic biology applications in ways that mimicked microblog discourse. We were interested in examining the following research questions: How do students respond to specific synthetic biology applications when limited to 140 characters in a microblog environment? How do students respond to peer's perspectives toward synthetic biology that were formed within such a constrained online setting?

## Background

### Socioscientific argumentation in science education

The Presidential Commission for the Study of Bioethical Issues recently released its first report calling for experts and policy makers to actively engage in public dialogue about the benefits and risks of synthetic biology in order for citizens to understand these effects on their fellow humans, nonhumans, and the environment as well as to share perspectives about the burgeoning field (Gutmann, 2011). As scientific knowledge is constructed by developing fields like synthetic biology, dispute and controversy emerge, so scientists engage in practices such as argumentation to resolve controversies and to reaffirm valid scientific claims. According to Driver, Newton, and Osborne (2000), new scientific knowledge does not become public knowledge until it has been vetted by the scientific community, and argument plays a crucial role in the practice of critical, public scrutiny. Similarly, the goals of science education are for students to develop the ability to evaluate the aims and results of those who produce scientific knowledge (Patronis, Potari, & Spiliotopoulos, 1999; Yoon, 2008a). Therefore, if science education is introducing students to claims produced by developing scientific fields, it should provide students opportunities to engage in the real-world argumentation practices carried out by scientists (Driver, Newton, and Osborne, 2000).

However, challenges associated with engaging students with contemporary SSIs have stemmed from the depiction of science as an uncontested and unproblematic body of knowledge that is free from critique (Driver, Leach, Millar, & Scott, 1996; Yoon, 2008b). Furthermore, while there have been curricular interventions that have improved students' abilities to evaluate SSIs, many have involved students interacting with texts and problem scenarios or their teachers (Sadler, 2004; Yoon, 2008b)—as opposed to experiencing these technologies first-hand. Incorporating synthetic biology into science education represents a paradigmatic shift from traditional science instruction. Instead of receiving scientific knowledge about bacteria or genetic engineering from an instructor, students genetically modify organisms and experience the field first-hand. As synthetic biology technologies move from university and corporate labs into K-12 education (Kafai et al., 2017), students will need to consider the benefits and risks of such biotechnologies, especially as they actively participate in a cutting-edge field that continues to evolve. In addition to providing hands-on learning opportunities, promoting discourse in which students share their ideas, negotiate multiple perspectives and claims, and evaluate new ideas enriches their learning (Yoon, 2008b). Given that students will participate in design through synthetic biology and engage with claims about this advancing field, classroom activities supporting their civic discourse through argumentation are much needed areas of education research.

### Microblogging, discourse and educational uses

Microblogs like Twitter are primarily used for everyday conversations, sharing information, and reporting current news (Java, Song, Finin, & Tseng, 2007). Due to the ability of posts to achieve high levels of interactivity and virality without geographical constraints, microblogs serve as significant platforms for various forms of communication, including political discourse (Colleoni, Rozza, & Arvidsson, 2014). However, the nature of such discourse on Twitter has been found to be highly partisan and worsening, necessitating a need to understand the social and technological dynamics central to this issue (Conover, Ratkiewicz, Francisco, Gonçalves, Flammini, & Menczer, 2011). Furthermore, even though Twitter exposes individuals to multiple, diverse perspectives, as a medium it may be insufficient for rational discourse and debate, as users tend to affiliate with views similar to their own (Yardi & boyd, 2010; Colleoni, Rozza, & Arvidsson, 2014). Therefore, there is a critical need to mitigate the limitations of discourse found in microblogs; one of the ways to do so is by examining the ways in which individuals engage in discourse through argument when microblogging about synthetic biology.

In educational contexts, microblogging can promote collaborative learning, encourage real-time reflective thinking, and increase student participation and engagement (Gao, Luo, & Zhang, 2012). When student microblogging occurred concurrently with classwide discussion at the college level, microblogging afforded new opportunities for discourse (Elavsky, Mislán, & Elavsky, 2011). However, the character limitation can impose constraints on students' critical thinking and self-reflection (Kassens-Noor, 2012). This limitation

can also affect the tone of the post, which can alter how readers perceive the post's content. Given that readers' interpretations of the potential risks of an emerging technology can be affected by a post's tone (Brossard & Scheufele, 2013), students need to be prepared to critically examine potential online claims about emerging technologies like synthetic biology by developing their skills in argumentation. While social networking sites have provided informal opportunities for students to engage in argument about SSIs (Greenhow, Gibbons, & Menzer, 2015), less research has attended to understanding how students engage in argument about synthetic biology on microblogging platforms.

In this study, we utilized an existing, commercial ARS called Mentimeter within our synthetic biology-based workshop that mimics the microblogging features of Twitter to support students' civic discourse through argumentation. Similar to Twitter, Mentimeter constrains responses to 140 characters. However, Mentimeter responses are anonymized and users are unable to interact with each other's posts by @-ing or retweeting. Students were responsible for evaluating multiple, anonymous perspectives and claims from their peers about specific synthetic biology applications. By facilitating a face-to-face discussion about these perspectives, we scaffold student engagement in the discursive practices of argumentation. Given that the discursive practices found in social media can inform students' collaborative discourse (Greenhow, Robelia, & Hughes, 2009), we examine how students form and evaluate claims about synthetic biology in a way that is relevant to their online communication.

## Methods

### Participants, context, and workshop

The SSI argumentation activity took place on the second day of a five-day, summer workshop that was part of a STEM program at a local science museum. The workshop participants included ten high school students who self-selected to participate in this workshop; however, nine students participated in this exploratory SSI workshop activity. Participants included five boys and four girls. The first two authors were both researchers and facilitators of the workshop, and both have had previously implemented and researched synthetic biology workshops with high school students (Kafai et al., 2017; Walker, Shaw, Kafai, & Lui, 2018).

The objective of the workshop was for participants to learn about and engage with synthetic biology. The workshop consisted of four main parts: 1) a wet lab where participants genetically transformed yeast, 2) mold design where participants had two opportunities to design molds for their cakes using silicone and various shapes, 3) recipe design where participants had two opportunities to bake their vitamin-enriched cakes using their own recipes, and 4) a pitch where participants presented their cakes to a panel of biologists and science educators. In addition to the laboratory and design activities, participants took part in an SSI argumentation activity where they discussed three different synthetic biology applications and the associated impact on humans, nonhumans, and the environment. The goal of the activity was to create an environment for students to share and evaluate multiple perspectives and claims about the applications. In particular, participants engaged in discourse about these applications using Mentimeter. A snapshot of the interface and responses are displayed in Figure 1.

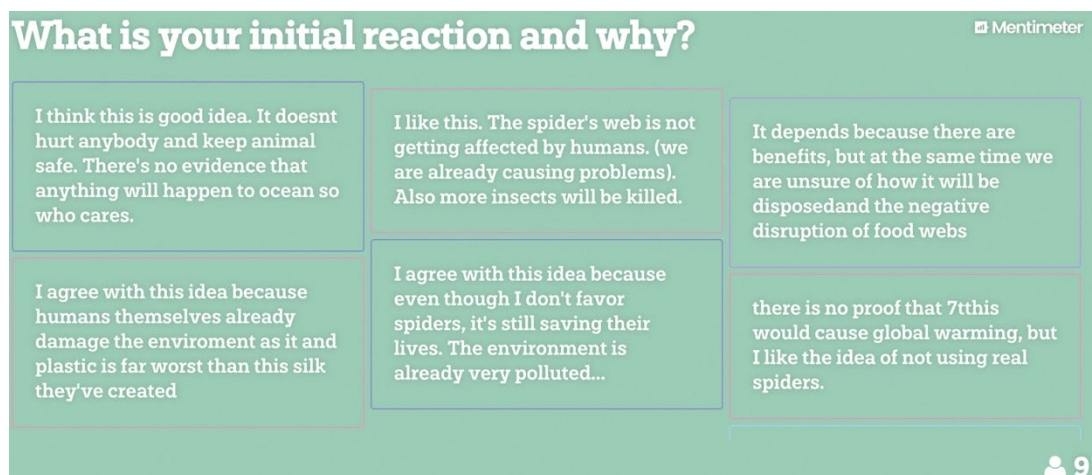


Figure 1. Screenshot of Mentimeter page displaying participant responses.

The entire microblogging activity took place for approximately 50 minutes. Participants read prompts explaining the benefits and risks of Golden Rice, CAR T-cell therapy, and Adidas Biofabric (see Table 1 for descriptions of each synthetic biology application and excerpts from the prompts). Using their individual computers, participants anonymously shared their initial reactions to each of the synthetic biology applications, which were immediately displayed for the entire class on the front screen to generate discussion. After reading everyone’s reactions, students reflected individually in response to two prompts: (1) After seeing the results, what do you find most interesting? Why? and (2) Find an argument that is most similar to yours and provide a possible counterargument. Lastly, participants engaged in a class-wide, face-to-face discussion based on their reflections.

**Table 1: Synthetic biology application**

| Application (type)            | Description   | Prompt excerpts   |
|-------------------------------|---|---|
| Golden Rice (plant microbial) | Rice is genetically-engineered to contain high levels of the vitamin-A-precursor, beta-carotene   | <i>“...Although Golden Rice has been available for some time now, it is unclear whether it has had the lasting benefits everyone expected. This is largely because Golden Rice tends to be more expensive to make than traditional rice. ...”</i> |
| CAR T-cell therapy (human)    | Newly developed cancer treatment that uses parts of the HIV virus to genetically modify white blood cells to recognize and destroy cancer cells | <i>“...Many experts in the field are calling this genetic engineering approach a new “pillar” of therapy because it does not involve rigorous chemotherapies or invasive surgeries....”</i>   |
| Adidas Biofabric (microbial)  | Shoes designed using synthetic spider silk  | <i>“...While this approach seems to be entirely beneficial, some critics have suggested that this will generate even more waste into the globe’s oceans, which are already plagued with record levels of plastic pollution...”</i>                |

### Data collection and analysis

Data collected for this study included the participants’ anonymized Mentimeter responses, individual worksheets participants used to reflect on peer responses, and video observations and fieldnotes. The first author began analysis by watching an excerpt of the video data of participant discussion of Golden Rice in order to observe how participants engaged in argumentation using Mentimeter. Through grounded analysis (Charmaz, 2002), the author noted and grouped participants’ discussion quotes in order to develop emergent, preliminary codes representing participants’ discursive practices when recording their initial evaluations about Golden Rice and critically examining their peers’ evaluations. These codes were evaluated and clarified through memoing and feedback from the second author. After we applied the codes to the remainder of the video observations and participant Mentimeter and worksheet responses, new themes emerged across the three data points. However, given that the focus of this is on the affordances of microblogging for engaging in argumentation, we decided to focus the remainder of the analysis on the Mentimeter and worksheet responses. Both authors revised the codes until consensus was reached and applied focused coding to the Mentimeter and worksheet responses by using the most significant and frequent codes. There were two categories of codes: among the initial Mentimeter responses, codes focused on participants’ attitudes toward each of the applications (e.g., not accept, balanced, accept, and unclear), and among the worksheet responses, codes focused on how participants evaluated peer claims (e.g. consensus, critiquing argument, justifying argument, etc.). Lastly, the authors drew relationships between these codes into themes that reflected how participant argumentation about synthetic biology was supported through the use of Mentimeter.

### Findings



## Finding 1: Participants’ evaluative claims depending on application type include a myriad of considerations

We begin by reporting on participants’ anonymous Mentimeter responses to three synthetic biology-based applications. We found that when participants were asked to describe their initial reactions toward the three applications using Mentimeter, they not only made evaluative claims but when their responses were aggregated and displayed to the whole class, two main narratives emerged. First, the type of application (e.g., microbial, human, and plant) influenced whether or not participants’ claims expressed support, lack of support, or a balanced perspective regarding the applications. Overall, the majority of participants found exclusively microbial applications acceptable (6 out of 8) while the majority of participants found human (6 out of 8) and plant applications (4 out of 8) unacceptable (see Table 2). Participants also made claims that were balanced, that is, neutral toward a microbial application overall. The following anonymous Mentimeter response reflects this as the participant noted, “it depends because there are benefits, but at the same time we are unsure of how it will be disposed and the negative disruption of food webs.” In making their claim, this participant did not express support or lack of support for Adidas Biofabric; instead, the student acknowledged the application’s benefits and risks. However, two participant responses were unclear but for disparate reasons. When evaluating Adidas Biofabric, it was challenging to determine one participant’s attitude toward the application as they explained, “there is no proof that this would cause global warming, but I like the idea of not using real spiders.” This participant’s use of the conjunction “but” makes it difficult to determine the tone of their claim. In addition, even though participants were tasked with writing one reaction, there was one extra anonymous response displayed that stated, “There is no proof that this will happen.” It is unclear what “this” the participant was referring to or if this response was intended as a continuation for their previous claim. These unclear examples allude to a potentially negative influence that a 140-character limitation can have on the tone of an argument, as well as how extraneous posts can distract from the coherent narratives found within a microblog’s feed.

Table 2: Range of attitudes found participants’ evaluative claims

| Application Type             | Support (N) | Not Support (N) | Balanced (N) | Unclear (N) | Total (N) |
|------------------------------|-------------|-----------------|--------------|-------------|-----------|
| Microbial - Adidas Biofabric | 6           | -               | 1            | 2           | 9         |
| Plant - Golden Rice          | 2           | 4               | 2            | -           | 8         |
| Human - CAR T-Cell           | 1           | 6               | 1            | -           | 8         |

The second narrative illustrates that in making their evaluative claims, participants made a wide range of considerations regarding the impacts of the synthetic biology application. When making their claims about Adidas Biofabric, participants frequently considered the specific impact of the microbial application on multiple contexts: humans, non-humans, and/or the environment. This is exemplified in one anonymous Mentimeter response as the participant explained, “I think this is good idea. It doesn’t hurt anybody and keep animal safe. There’s no evidence that anything will happen to ocean so who cares.” They not only expressed support for the Adidas Biofabric but considered the impact of the microbial application on humans (“anybody”), non-humans (“animal”), and the environment (“ocean”). This was the only response that considered the impact of the application on all three contexts. When making their claims about Golden Rice and CAR T-cell therapy, participants considered the impact of the plant-microbial and human applications in broad terms and/or to humans in particular. This point was illustrated when an anonymous Mentimeter response disapproved of CAR T-cell therapy as the participant noted, “NO! It could easily attack other tissues and foreign thing that could be good for you. It could be very dangerous if it gets out of control.” This participant cautioned that the potential impact on humans makes this application of synthetic biology unacceptable. However, another participant made broad considerations when evaluating Golden Rice, declaring, “My initial reaction was rather between the two arguments. There is the chance of both negative and positive outcomes, so anything can happen.” Even though their claim was of a balanced perspective, it lacks specific considerations of Golden Rice’s impact, making it a challenge to determine whether or not this participant fully understood the application’s benefits and risks. Finally, participants also based their considerations on benefits, costs, and usefulness of the application. In

response to Golden Rice, an anonymous Mentimeter response asserted, “in between. It is very expensive to make and it could be harder to harvest. However most communities that are Vitiam A deficient could use.” Here, the participant explained that costs and usefulness are central considerations in their overall claim, which is balanced. These findings suggest that microblogging has the potential of supporting students’ critical thinking regarding the impacts of synthetic biology applications.

## **Finding 2: Participants’ argument-based rationales evaluate other perspectives.**

After displaying all of the participants’ 140-character, anonymous Mentimeter responses, participants answered the following question on individual worksheets: “After seeing the results, what do you find most interesting? Why?” We found that when participants read all of their peers’ perspectives on Mentimeter, they engaged a claim in terms of whether or not it aligned with their own and evaluated it according to the rationale presented in the claim. Regarding the three synthetic biology applications, participants frequently asserted agreement with a given rationale: microbial applications (4 of 8 responses), plant-microbial applications (4 of 7 responses), and human applications (4 of 8 responses). This was exemplified in responses like Natalie’s when she explained, “the most interesting reasoning is the one that speaks about the [effects] on the environment but also the potential good things that come out of it. This is because it [exhibits] the potential outcome on both sides.” Like several participants, Natalie asserted her agreement with this claim regarding Adidas Biofabric based on the fact that the rationale is balanced and in consideration of two viewpoints. In addition to asserting agreement, argument building was reflected in moments when students indicated they agreed with a given claim and advanced the rationale by asserting additional ideas in support of the given perspective. This was evident in Edward’s point when he explained, “I found the ‘sold at higher’ price one interesting I say this because you can’t simply sell something that you aren’t at complete assurance of it being harmless.” Here, Edward adds to a peer’s lack of support towards Golden Rice being sold at higher prices without knowing the benefits by asserting the need for assurance that the application is harmless. The frequent asserting of agreement and argument building alludes to the potential of microblogging in reinforcing participants affiliating with peers who share their views.

Participants also asserted critiques towards given claims and their rationales. When participants considered microbial-based perspectives that were contrary to their own, their evaluations were also rationale-based and included a critique of the opposing perspective. This was illustrated in participant responses like Michael’s who explained, “I think the response that says that we should use it because we already damage the environment is interesting. The reasoning is kind of dumb. It’s like saying if a car is broke, you should break it some more.” Michael pointed out that the rationale to use synthetic biology-based applications on plants despite the potential environmental effects because the environment is already polluted is erroneous and uses an analogy to underscore his point. Even though only two participants critiqued their peers’ claims and rationales, this demonstrates the potential of participants being able to critically evaluate claims about synthetic biology they may be exposed to in microblogging platforms.

Instances when participants pointed to a perspective and its rationale as illuminating occurred infrequently. These perspectives reflect times when students found a perspective and rationale different from their own, but interesting nonetheless as Carla noted, “I found the fact that other people mentioning the price was interesting because I personally didn’t consider the price.” As Carla explained, the rationale advanced by one of her peers was a perspective she had not considered previously but was something she found notable. This supports the potential of microblogs in exposing participants to perspectives and discussions they may not have been witnessed before.

## **Finding 3: Participants weigh counter arguments in relation to their own**

Lastly, when answering the worksheet question: “Find an argument that is most similar to yours and provide a possible COUNTER ARGUMENT,” participants were able to develop counter arguments to claims aligned with their own as displayed on Mentimeter. This occurred across synthetic biology-based contexts including microbial applications (3 of 7 responses), plant-microbial applications (4 of 7 responses), and human applications (3 of 7 responses). For instance, when Heather considered her own claim about using synthetic biology on microbes, she noted, “although there are many benefits there is a possibility to hurt food webs which is a major issue.” Heather weighed the potential benefits that guided her overall counter argument against the potential impacts the technology could have on the environment. With regard to plants, a similar outcome occurred as illustrated in Leah’s point that “golden rice... tastes good and has enough Vitamin A; although, no one knows the impact which could be a bad thing.” Here, Leah broadly considered the potential for adverse outcomes when using synthetic biology in plant-based applications.

Similarly, students weighed counterarguments against their own perspectives about human-based synthetic biology applications. This is exemplified in Edward’s response as he explained, “I’m going to use the

point from earlier. The counter argument could be: you have to give it a chance, because you could say the same thing about most inventions. You don't know the risks but you have to take it." Here, Edward noted that despite the potential harm that could arise from using synthetic biology on human cells—a perspective that aligns with his own—there are risks associated with most technological advances, making this not a unique situation as well as a necessary part of making progress. Participants' ability to construct counter arguments based on their peer's 140-character claims alludes to the potential of considering perspectives they may be exposed to on microblogging platforms that conflict with their own.

## Discussion

In this study, we facilitated a discussion about synthetic biology by using an online microblogging platform which, like Twitter, limits responses to 140 characters and by observing how students engage in argument through such constraints. Despite the character limitation, students were able to construct concise yet diverse evaluations of the three different applications based on a range of considerations (impact on humans, environment, and animals and their usefulness). In addition, students were able to form and weigh counterarguments against their own initial arguments. Previous research suggests that Twitter can foster an increased creating and sharing of ideas beyond the classroom due to the automatic tracking of tweets by day and time, supporting continuous participation of users (Kassens-Noor, 2012). Even though Mentimeter lacks this ability to track responses across time, our findings suggest that such platforms can foster combined knowledge *within* the classroom due to participants' ability to examine and draw arguments from multiple peer perspectives. Furthermore, even though Twitter provides identifying information about users, Mentimeter anonymizes users. Given that anonymity can improve discussion participation (Hsi & Hoadley, 1997), its use may have mitigated the social influence effects found on social media, allowing more students' evaluative claims to be drawn upon.

However, considering that most students did not make considerations about the synthetic biology applications simultaneously across humans, nonhumans, and the environment, the 140-character limitation may have constrained their ability to do so. When presenting the scenario for each application, we provided potential impacts of the application on humans, nonhumans, and the environment. Nonetheless, participants' arguments tended to be based on impacts that were either broad or concerning just one or two of the three impacts we provided. Considering that evaluating the impacts across all these categories is a quintessential part of being a steward of humans, nonhumans, and the environment as a whole (Gutmann, 2011), this observation illustrates a potential limitation of microblogging in that 140 characters may have forced participants to focus on a topic without allowing them to express complex thoughts (Junco, Heiberger, & Loken, 2011). In addition, it is interesting that in evaluating their peers' perspectives, the majority of students expressed agreement with their peers' arguments. Given the complexity of the issues surrounding synthetic biology, it is possible students may be drawn to perspectives that are similar to their own, reinforcing an echo chamber (Colleoni, Rozza, & Arvidsson, 2014) of their own bias and solidifying them on these platforms.

While this study was exploratory in nature, it did illuminate some interesting phenomenon that should be explored in further research. An element of SNSs that was not present in our discussion was the ability for students to respond to their peers' perspectives online as well as read each other's response. Given that microblogging can document learning processes in real-time, students themselves could better understand their own and their peers' argumentation process given the proper scaffolds. This documentation of their argumentation could be utilized to highlight the complexity of synthetic biology. This study provides a start to exploring the affordances and constraints of microblogging features such as character limitations in engaging student in argumentation surrounding synthetic biology and other SSIs.

## References

- Auer, M., Zhang, Y. & Lee, P. (2014). The potential of microblogs for the study of public perceptions of climate change. *Wiley Interdisciplinary Reviews: Climate Change*, 5(3), 291-296.
- Brossard, D. (2013). New media landscapes and the science information consumer. *Proceedings of the National Academy of Sciences*, 110(3), 14096-14101.
- Brossard, D. & Scheufele, D. (2013). Science, new media, and the public. *Science*, 339(6115), 40-41.
- Charmaz, K. (2002). Stories and silences: Disclosures and self in chronic illness. *Qualitative Inquiry*, 8(3), 302-328.
- Cheng A. & Lu, T. (2012). Synthetic biology: An emerging engineering discipline. *Annual Review of Biomedical Engineering*, 16(14), 158-178.

- Colleoni, E., Rozza, A. & Arvidsson, A. (2014). Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. *Journal of Communication*, 64(2), 317-332.
- Conover, M.D., Ratkiewicz, J., Francisco, M., Gonçalves, B., Flammini, A. & Menczer, F. (2011). Political polarization on Twitter. In Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (p. 89-96). Barcelona, Spain: AAAI.
- Driver, R., Leach, J., Millar, R., & Scott, P. (1996). Young people's images of science. Philadelphia, PA: Open University Press.
- Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education*, 84, 287-312.
- Elavsky, C., Mislan, C., & Elavsky, S. (2011). When talking less is more: Exploring outcomes of Twitter usage in the large-lecture hall. *Learning, Media, and Technology*, 36(3), 215-233.
- Gao, F., Luo, T. & Zhang, K. (2012). Tweeting for learning: A critical analysis of research on microblogging in education published in 2008-2011. *British Journal of Educational Technology*, 43(5), 783-801.
- Greenhow, C., Robelia, B., & Hughes, J. (2009). Learning, teaching, and scholarship in a digital age: Web 2.0 and classroom research: What path should we take now? *Educational Researcher*, 38(4), 233-245.
- Greenhow, C. & Gleason, B. (2012). Twitteracy: Tweeting as a new literacy practice. *The Educational Forum*, 76(4), 464-478.
- Greenhow, C., Gibbins, T., & Menzer, M. M. (2015). Re-thinking scientific literacy out-of-school: Arguing science issues in a niche Facebook application. *Computers in Human Behavior*, 53, 593-604.
- Gutmann, A. (2011). The ethics of synthetic biology: Guiding principles for emerging technologies. *The Hastings Center Report*, 41(4), 17-22.
- Hsi, S. & Hoadley, C. (1997). Productive discussion in science: Gender equity through electronic discourse. *Journal of Science Education and Technology*, 6(1), 23-36.
- Java, A., Song, X., Finin, T. & Tseng, B. (2007). Why we Twitter: Understanding microblogging usage and communities.
- Junco, R., Heiberger, G., & Loken, E. (2011). The effect of Twitter on college student engagement and grades. *Journal of Computer Assisted Learning*, 27(2), 119-132.
- Kafai, Y., Telhan, O., Hogan, K., Lui, D., Anderson, E., Walker, J., & Hanna, S. (2017). Growing designs with biomakerlab in high school classrooms. In Proceedings of the 2017 Conference on Interaction Design and Children (pp. 503-508). New York, NY: ACM.
- Kassens-Noor, E. (2012). Twitter as a teaching practice to enhance active and informal learning in higher education: The case of sustainable tweets. *Active Learning in Higher Education*, 13(1), 9-21.
- Murphy, D. (2013). *Twitter: Digital Media and Society Series*. Cambridge, UK: Polity Press.
- Patronis, T., Potari, D., Spiliotopoulos, V. (1999). Students' argumentation in decision-making on a socio-scientific issue: Implications for teaching. *International Journal of Science Education*, 21(7), 699-702.
- Shäfer, M. (2012). Online communication on climate change and climate politics: a literature review. *WIREs Climate Change*, 3(1), 527-543.
- Walker, J., Shaw, M., Kafai, Y., & Lui, D. (2018). Biohacking food: A case study of science inquiry and design reflections about a synthetic biology high school workshop. In Proceedings of the Thirteenth International Conference of the Learning Sciences (p.1559). London, UK: ISLS.
- Yardi, S. & boyd, D. (2010). Dynamic debates: An analysis of group polarization over time on Twitter. *Bulletin of Science, Technology & Society*, 30(5), 316-327.
- Yoon, S. (2008a). An evolutionary approach to harnessing complex systems thinking in the science and technology classroom. *International Journal of Science Education*, 30(1), 1-32.
- Yoon, S. (2008b). Using memes and memetic processes to explain social and conceptual influences on student understanding about complex socio-scientific issues. *Journal of Research in Science Teaching*, 45(8), 900-921.

## Acknowledgements

This work was supported by a grant (#1623018) from the National Science Foundation to Yasmin Kafai and Orkan Telhan. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation or the University of Pennsylvania. The authors wish to thank Debora Lui, Gayithri Jayathirtha, and Joyce Yu for their help with developing and supporting workshop activities.

# Developing Productive Discourse Through Collective Inquiry of Knowledge-Building Principles

Yuyao Tong, The University of Hong Kong, tongyuyao2016@gmail.com  
Carol K. K. Chan, The University of Hong Kong, ckkchan@hku.hk

**Abstract:** This study investigates how secondary school students inquired into knowledge-building principles and engaged in meta-talk for productive knowledge building. The study involved a class of Grade 10 students in Hong Kong, working on collective inquiry in Art and Design using Knowledge Forum® (KF), and reflecting on their discourse, using knowledge-building principles. A key design involved students writing on KF inquiring into knowledge-building principles and meta-talk on the processes. Results indicated that the students engaged in productive online discourse on both domain and discourse understanding progressively. Classroom analyses suggested how students engaged in meta-talk and inquiry into principles with corresponding change in the knowledge about discourse. The implications of scaffolding students' inquiry of principles in a knowledge-building environment are discussed.

## Introduction

Dialogue, as a talk that combines the sharing of ideas with meaning making, acts as a medium for developing students' learning and thinking (Mercer & Littleton, 2007) and specifically important for improving ideas and creating new knowledge (Bereiter & Scardamalia, 2016). It provides a theoretical framework for computer-supported collaborative learning (CSCL) (Stahl, Cress, Ludvigsen, & Law, 2014). Knowledge building (KB), as a CSCL model, aims to advance community collective knowledge via idea-centered discourse (Scardamalia & Bereiter, 2014). Scardamalia and Bereiter (2014) argued that knowledge building will fail if knowledge-building dialogue fails; therefore, the essential goal in a knowledge-building environment is to help students take collective responsibility for developing knowledge-building dialogue. To support such dialogue, Knowledge Forum® (KF) has been designed as a collaborative discourse space in which students can develop discourse from a knowledge-building perspective, with the aid of its various technological functions (e.g., build-on others' ideas, linking and rise-above ideas, and revise ideas, etc.) and metacognitive scaffolds (e.g., "I need to understand", "My theory", "A better theory"). Further, Scardamalia (2002) proposed twelve principles to characterize knowledge building and to guide teachers and students to engage in a principle-based instead of procedure-based learning environment; these principles include epistemic agency, improvable ideas, community knowledge, idea diversity and so forth that are the essential themes of knowledge building. While it is a taken-for-granted theme but limited systematic study of what such understanding of principles would influence students' productive discourse engagement and how to scaffold such understanding is not well examined. Many KB studies assert the importance of principles (van Aalst & Chan, 2007; Zhang et al., 2018), but this is one of the few to develop a systematic design focusing on students' understanding and collective inquiry of KB principles based on their authentic practice and work thereon rather on their their declarative knowledge of a list of principles.

Examining principles as conceptual and epistemic artefacts for inquiry is critical for knowledge work as they are the epistemic criteria for knowledge generation. Many studies have been analyzed and scaffolded productive KF discourse (Chan, Lam, & Leung, 2012; Niu & van Aalst, 2009; Zhang, Scardamalia, Lamon, Messina, & Reeve, 2007), but how students understand the nature of discourse and specifically, how it needs to be linked to KB principles and its influence on students' productive discourse engagement has not been understood. In related research, Wegerif (2001) developed dialogic pedagogy for teaching of dialogic skills and ground rules (e.g., all relevant information is shared), to help students develop meaningful discourse for problem solving. Similar to the teaching of ground rules to provide students' explicit knowledge about dialogue, we propose that students need to have an understanding of the nature of discourse; students' epistemic understanding of knowledge building and principles would be important in students' creative work for productive discourse. This study addresses this issue, as part of a large study that examining the design, process, and roles of a meta-talk computer-supported knowledge-building environment for students' understanding of discourse and knowledge advance. In a preliminary study, we reported that students' understanding of discourse could influence their taking collective responsibility in the community discussion. This study examines further on how students' collective inquiry of principles can be scaffold and its relation to knowledge-building dialogue engagement, especially for low-achieving students who might have difficulties in engaging in higher-order thinking skills (Zohar et al., 2001) and have less opportunity to take responsibility to engage in productive

discourse as teachers believe that low-achieving students need more direct instruction (Zohar & Dori, 2003). Overall, this study designed an environment and examined the role of design on students' productive KF discourse and increased collective responsibility; how students reflected on their discourse and engaged in meta-talk and build knowledge about KB principles; and the influences of these processes on their subsequent understanding on knowledge building. Specifically, three research questions were addressed: (1) Did students engage in productive KF discourse with increasing collective cognitive responsibility over time? (2) How did the dynamics of classroom meta-talk and processes scaffold students' collective inquiry of principles? And (3) How was students' understanding of the nature of discourse reflected in interviews and what were its relations to the KF knowledge-building dialogue engagement?

## Methods

### Design the knowledge building meta-talk environment: collective inquiry of principles

This study drawn upon the data from students' KF discourse, classroom talk, and interviews involving twenty-one Grade 10 students from a low-performing school based on students' public examination results studying the topic on "Green Design" participated in the study over one semester. The key design was to include different KF views and one specifically devoted to inquiry into KB principles. Specific designs included: (1) KB Wall and ideas development trajectory (students posted and shared their initial ideas on a wall with others' build-on, followed by a classroom reflection) (Weeks 1-3); (2) authentic problems and KF inquiry (students proposed authentic problems by watching videos on "design and earth environment" and group discussion, followed by KF inquiry) (Weeks 4-6); (3) deepening inquiry through comparing discourse (students worked in groups to generate knowledge about their prior understanding of discourse by comparing their KF discussions) (Week 7-9); (4) connecting students' understanding and KB principles (students connected and matched their prior understanding on discourse with the epistemic criteria for discourse that are the KB principles) (Weeks 10-12); and (5) collective inquiry and reflection (students wrote portfolio notes on KF to reflect on their collective understanding of principles) (Weeks 13-15).

## Data Analysis and Results

### RQ1. Examining KF engagement and inquiry into principles and domain knowledge

#### Students' KF engagement and change towards connectivity and collective responsibility

The first research question examined student engagement and inquiry into domain knowledge and principles and change over time. Primarily, the key question is can students engage in productive discourse? We first conducted KBDex (Oshima et al., 2012) analysis to explore how discourse network changed between the Period 1 (before classroom meta-talk on principles - Week 1 to 6) and Period 2 (after classroom meta-talk on principles - Week 7 to 15). Students' KF notes were exported into KBDex, which produced three networks of analysis-students, discourse, and keywords-based on the selected conceptual keywords. We examined the discourse network change. As Figure 1 shows, the discourse network is more integrated with few fragmented notes in Period 2 than Period 1. In Period 1 (1a and 1c), there are many separate notes (red) remained outside the main cluster (yellow). In Period 2 (1b and 1d), the discourse network was more integrated, with fewer fragmented notes (red). This suggests that in Period 2, students engaged in KF discourse in a more cohesive discussion and productively with sustained collective knowledge advancement than Period 1.

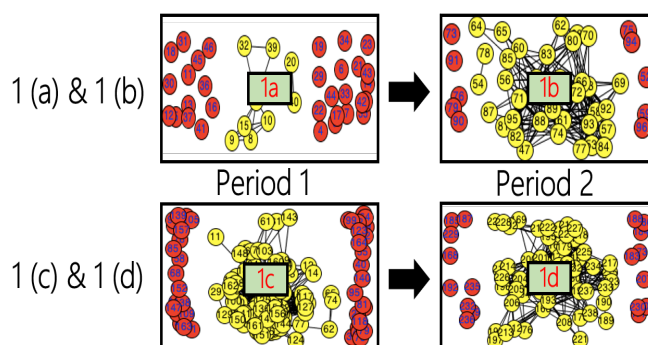


Figure 1. KF discourse network change - principles (1a&1b) and domain discussion (1c&1d) in two periods.

### Inquiry thread analysis of online KF discourse

Further to overall network analyses, content analysis was conducted to examine how students engaged in productive knowledge building. The KF notes were parsed into threads based on the conceptual problems (Zhang et al., 2009), with 14 such threads being identified (e.g. “design and environmental development”). Individual notes were coded within each inquiry thread, using a coding scheme based on a theory- and data-driven approach and including questions (Hakkarainen, 2003), contributions (Chuy et al., 2011), and meta-discourse (Table 1). As Table 2 shows, the results suggested that students asked sustained questions for deeper inquiry, enriched and improved on ideas, and used elaboration to support those ideas; in addition, students also engaged in meta-discourse by referring back to previous discussions to synthesize what had been discussed and rise-above the inquiry process using reference function incorporating other community ideas. A second rater coded 30% of data, and the inter-rater reliability was .92 for questioning, .86 for theorizing, and .87 for community (Cohen’s kappas).

**Table 1: Coding scheme for analyzing KF discourse in inquiry threads**

| Codes                            | Sub-codes                 | Description   | Examples   |
|----------------------------------|---------------------------|---|--|
| Questioning and Identifying Gaps | Fact-seeking              | Questions on seeking factual information  | What is the definition of design?  |
|                                  | Explanation-seeking       | Questions on seeking open-ended responses with explanation  | How to use design to solve environmental problems?   |
|                                  | Sustained inquiry         | Asking further questions based on previous notes or ideas and make the discussion deeper  | How can a design be regarded as a good design if it cannot solve any problems? (A question asked based on a previous note).  |
| Theorizing and Improvable Ideas  | Simple claim              | Simple (dis)agree or repeat a statement   | I agree with the idea.   |
|                                  | Proposing an explanation  | Proposing a theory that explain certain phenomena for the first time  | Design can improve the quality of our life.  |
|                                  | Enriching an explanation  | Enriching a theory with elaboration and new information   | A good discourse needs to have a good question first...The question can be solved through discussion and new questions would emerge...   |
|                                  | Supporting an explanation | Supporting an already existing theory proposed by another student and providing a justification   | It could be or not, but we can learn new knowledge and improve ideas through communication.  |
|                                  | Improving an explanation  | Improving an already existing theory through elaboration, specifying details and using new evidence   | A good discourse refers to a community discussion rather than individuals expressing ideas. Students can build-on other’s ideas...we can acquire valuable ideas through discussion.  |
|                                  | Challenging               | Challenging or disagreement an existing ideas proposed by other students  | What you mentioned about the use of design...but I do not think we can solve all the problems by our artifacts design.   |
| Meta-discourse                   | Connection                | Reference to their own or others’ notes, or quoting extra sources to advance understanding  | A good discourse needs to have diverse ideas, as proposed by [red] (“diverse ideas for discourse” S077)...we also need to develop a productive discourse with meaningful ideas.  |
|                                  | Rise-above                | Refer back to what has been discussed and ask a new question for monitoring the inquiry process and advancing the discussion; or generating an explanation or evaluation to appraise their own or others’ notes | Through our discussion, we discussed the idea of design...by [red] (“design” student S089) and design can be developed together with the society by [red] (“design and development” student S005) What we need to discuss further is how can the design be improved? |
|                                  | Synthesize                | Coordinating and integrating collective ideas from previous discussion with a reflection  | [red] (“a good discourse needs to have diverse ideas...”) by student084 [red] (“...we need to querying others’ ideas”) by student011. In this discussion, they think that students need to participate actively with diverse ideas...                                |

Table 2: Number of different codes in inquiry threads

| Thread | Fact | Explain | Sustain | Simple | Propose | Enrich | Support | Improve | Connect | Rise | Synthesize |
|--------|------|---------|---------|--------|---------|--------|---------|---------|---------|------|------------|
| #1     | 0    | 5       | 7       | 4      | 18      | 6      | 1       | 1       | 3       | 1    | 0          |
| #2     | 0    | 4       | 5       | 4      | 14      | 5      | 4       | 3       | 7       | 2    | 2          |
| #3     | 1    | 4       | 8       | 2      | 5       | 7      | 3       | 2       | 3       | 2    | 0          |
| #4     | 1    | 2       | 12      | 2      | 19      | 7      | 2       | 0       | 8       | 1    | 0          |
| #5     | 0    | 2       | 9       | 2      | 5       | 1      | 3       | 4       | 4       | 2    | 4          |
| #6     | 0    | 4       | 1       | 1      | 4       | 0      | 1       | 0       | 0       | 0    | 0          |
| #7     | 0    | 4       | 3       | 0      | 5       | 1      | 2       | 2       | 2       | 2    | 2          |
| #8     | 0    | 3       | 5       | 3      | 7       | 4      | 2       | 4       | 3       | 0    | 1          |
| #9     | 0    | 0       | 1       | 0      | 4       | 4      | 9       | 0       | 0       | 2    | 3          |
| #10    | 1    | 0       | 1       | 0      | 2       | 1      | 0       | 1       | 0       | 0    | 0          |
| #11    | 0    | 1       | 5       | 2      | 9       | 1      | 0       | 3       | 1       | 1    | 1          |
| #12    | 0    | 2       | 6       | 0      | 6       | 9      | 3       | 4       | 10      | 2    | 8          |
| #13    | 0    | 2       | 3       | 0      | 1       | 2      | 2       | 0       | 0       | 0    | 0          |
| #14    | 0    | 3       | 2       | 0      | 8       | 10     | 0       | 1       | 0       | 0    | 3          |
| Total  | 3    | 36      | 68      | 20     | 106     | 58     | 33      | 25      | 41      | 15   | 24         |
| Mean   | 0.21 | 2.57    | 4.86    | 1.54   | 7.57    | 4.14   | 2.36    | 1.79    | 2.93    | 1.07 | 1.71       |
| SD     | 0.41 | 1.49    | 3.2     | 1.45   | 5.37    | 3.16   | 2.16    | 1.52    | 3.17    | 0.88 | 2.19       |

To understand how students' discourse changed, the KF notes were divided into two periods-before and after the classroom meta-talk on KB principles. We calculated the frequency of high-level discourse moves (Table 3), and analyses were conducted to examine whether students engaged in more productive discourse from Period 1 to Period 2. Paired sample t-test indicated that there were significant increases from Period 1 to Period 2 in terms of sustained inquiry,  $t(20) = 2.597, p < .05$ ; enriching an explanation,  $t(20) = 3.525, p < .01$ ; supporting an explanation,  $t(20) = 3.068, p < .01$ ; improving an explanation,  $t(20) = 2.958, p < .01$ ; connection,  $t(20) = 3.118, p < .01$ ; rise-above,  $t(20) = 2.366, p < .05$ ; and synthesize,  $t(20) = 2.915, p < .01$ . The results suggest students' discourse moves changed to a more knowledge-building approach, which the students engaged more in supporting and improving their ideas and explanation. Further, students also engaged in a reflection approach with collective ideas coordinating and monitoring collective ideas for deeper inquiry. Overall, the results showed that students engaged in a more productive knowledge-building discourse over time.

Table 3: Frequency of high-level notes in Period 1 and Period 2

| High-level discourse         | Period 1 (Week 1 to Week 6) | Period 2 (Week 7 to Week 15) |
|------------------------------|-----------------------------|------------------------------|
| Explanation-seeking question | 20                          | 16                           |
| Sustained inquiry            | 29                          | 39                           |
| Enriching an explanation     | 28                          | 30                           |
| Supporting an explanation    | 8                           | 25                           |
| Improving an explanation     | 2                           | 23                           |
| Challenging                  | 4                           | 10                           |
| Connection                   | 9                           | 32                           |
| Rise-above                   | 0                           | 15                           |
| Synthesizing                 | 0                           | 24                           |

## RQ2. How did students engage in meta-talk for collective inquiry of principles?

This study also analyzed themes identified from the classroom processes to investigate how students could engage in productive KF discourse (RQ1). We discussed the design and examined how students engaged in collective inquiry of principles mediated by classroom meta-talk for knowledge building.

### Scaffolding principle-based understanding with KB Wall idea reflection and KF inquiry

Generating productive discourse is key to knowledge building for community knowledge advancement. In the knowledge-building classroom, scaffolding is initially used to help students engage in a principle-based collaborative learning environment. The knowledge-building wall (KB Wall) (Fig 2a), as a physical visualization of KF, provides an opportunity for students to publish their ideas by posting their questions, and to build-on each other's ideas (the lines represent the build-on relations). Further, to scaffold students to track their idea development on the KB Wall, students were asked to write rise-above portfolios to reflect on how their



ideas improved with referencing the community's ideas from the KB Wall (Figures 2b and 2c). Students shared their rise-above portfolio and refined their ideas through classroom discussion while continuing their discussions on KF (Fig 2d), both in terms of domain knowledge and understanding of principles. KF provided a collaborative space for improving ideas and sustaining inquiry through idea-centered discourse. This KB Wall and rise-above portfolio writing experience might help students engage in the sense of principles of improvable ideas, rise-above, epistemic agency, and community knowledge.

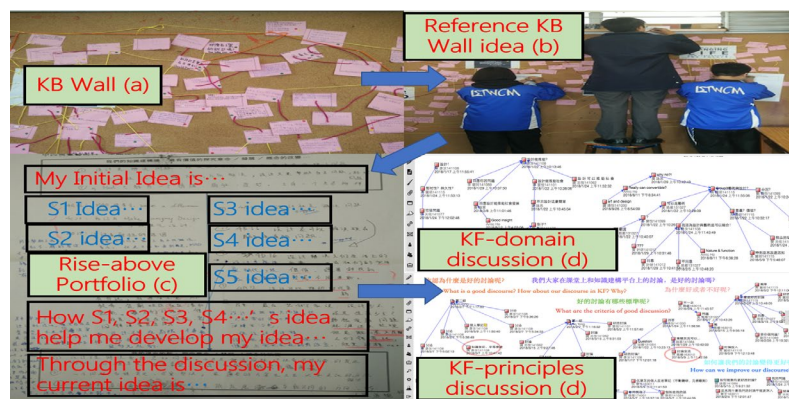


Figure 2. KB Wall (a) - Rise-above portfolio (b&c) - KF inquiry (d).

### Comparing KF discourse for developing criteria of knowledge building discourse

Understanding the knowledge-building principles, which are the epistemic criteria for the nature of discourse and standards of knowledge work is important for producing knowledge-building dialogue. After discussing matters on KF for a period of time, students started to reflect on their discourse, using criteria for developing of a good discourse. Two discourse moves that students generated in KF were provided (Fig 3a), along with a designed prompt sheet (Fig 3b). Students drew mind map, which is one of the scaffold to help low-achieving students to visualize their ideas, as a way to explain their comparison and understanding of the discourse moves (Fig 3c). They shared their mind map with the class, followed by a classroom discussion. The following excerpt illustrates how criteria for good discourse were developed by students through their KF discourse reflection. The discussion suggests that these students were not only described the structures of the threads, they also assessed the quality of the notes (*they have a summary note in the discourse move*) and noted the strategies for improving a discourse (*we can make a summary to reflect...and integrate the different ideas together*).

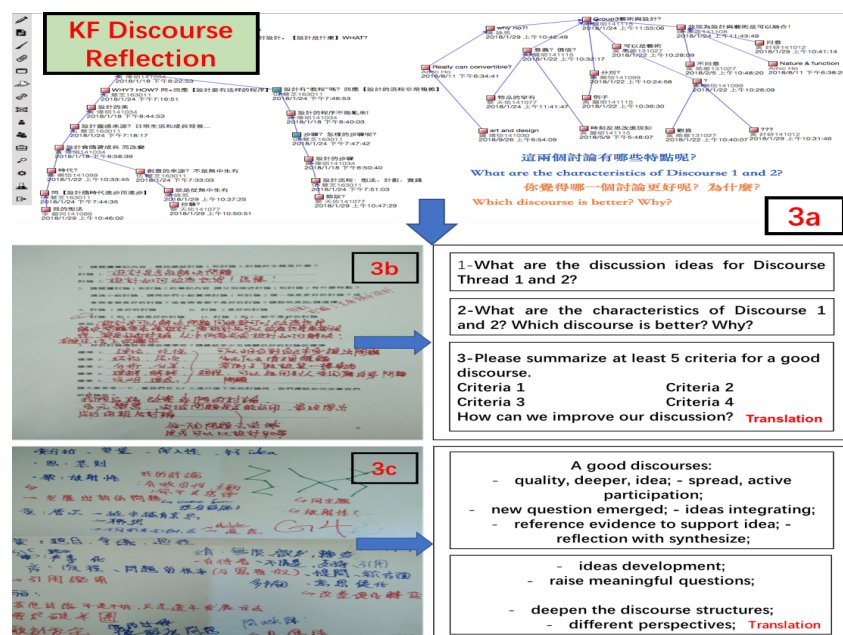


Figure 3. Prompt sheet for criteria generating (3b) and Mind map on discourse comparison (3c).

- T(Teacher) Can anyone explain your understanding of a good discourse based on your community KF discussion comparison?
- S7(Student) We compared two discourse moves that we proposed in KF...The first discourse move was more **singular with a few straight build-on lines**...The other discourse move was **more diversified with spread build-on lines**...their **idea was getting deeper and had the potential to produce a new question to sustain the inquiry**...
- T Good point! Can anyone add more?
- S7 We can **make a summary to reflect on what we had discussed**...
- S8 ...The **“diverse” refers to the idea development**...
- S9 They had a summary note in the discourse move; the **“summary” was one of the criteria we proposed for a good discourse**...

### Identifying and categorizing epistemic criteria linked to principles

Another aspect of scaffolding students’ understanding of principles is the linking the criteria generated by students (learner criteria) and the knowledge-building principles (expert criteria). Students constructed mind maps that categorized the learner criteria into different groups and linked them to “expert criteria” (Fig 4). For example, they put the “reference others’ idea” and “synthesizing and reflection” into one group. Next, they started to link these learner criteria to KB principles and found some of the principles were overlapped different groups. Different arrows are connected the criteria generated by students and the KB principles. Students were asked to explain the reasons for their categorizing and linking. The teacher invited students to explain the categories and links they developed and the following example showed how students explain the reasons (*reflection is a type of assessment that can help us to revise and improve ideas*...) and gradually engaged in a deeper discussion in illustrating how to improve a discourse with using these principles as strategies.

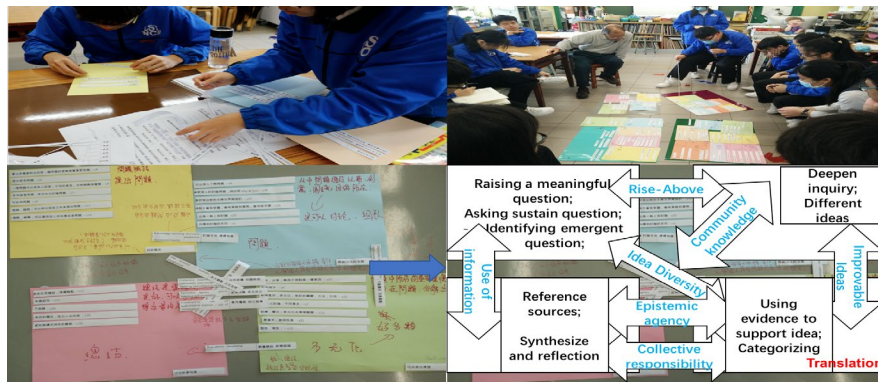


Figure 4. Criteria categorizing and principles matching and Mind map with translation.

- T Can anyone explain how the connections you made between your criteria and the principles?
- S1 We put the principle **“improvable ideas”** here because **the question in this discourse move made the ideas developed**... ..
- S3 Yes, **reflection is a type of assessment that can help us to revise and improve ideas in the discussion**.
- S1 For the **“collective responsibility”**, we connected it the **contribution**.
- T Can you explain a bit more?
- S4 It refers to that everyone should take the responsibility **to join the discussion and contribute to the community**...

### Collective inquiry with portfolio notes on principles understanding

Students continued to work on KF and wrote portfolio notes reflecting on their understanding on the nature of good discourse and linking it to KB principles. Figure 5 shows an example of KF view for collective inquiry on principles and a portfolio note using reference function using to explain their understanding.

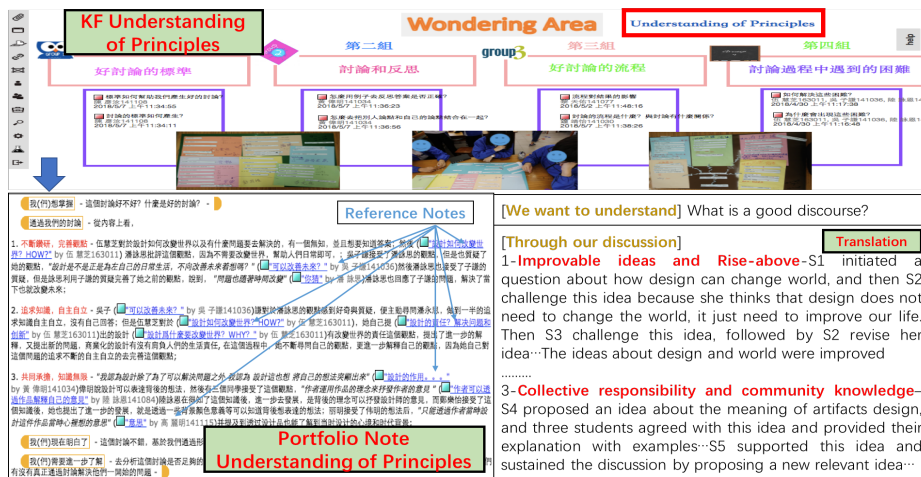


Figure 5. Portfolio note for reflection on understanding of discourse with translation.

### RQ3. Deepening understanding of principles and relations with KF engagement

The third research question examined the role of design and processes on students' understanding of discourse and the shift towards principles as epistemic criteria. We examined students' open-ended questions and interviews and we also examined how their understanding was related to their productive KF discourse.

#### From viewing good discourse as acquiring an agreement to viewing it as improving ideas

Qualitative analysis of students' interviews showed that the changes in their understanding of discourse between their pretest and posttest interviews. Students initially thought the purpose of discourse was only to get an agreement or a correct answer, after the program, their understanding of discourse changed to reflect a KB approach. Student LYE thought the major theme of discussion was to get answers in the pretest; however, in the posttest, she started to think that the purpose of discourse was to improve ideas through sustained discussion.

|                    |  |
|--------------------|--|
| Pretest Interview  | "A good discussion is asking a question together...to get answers through discussion."   |
| Posttest Interview | "A good discussion <i>connects ideas together for deeper discussion</i> ...The initial question for this cluster of discussion was "what is design?"...The <i>idea about design continued to improve through the discussion</i> ." |

#### From viewing good discourse as a process of merely discussion to viewing it as rise-above

To engage in a productive discourse, students need to querying ideas, coordinate ideas with reflection and rise-above for further inquiry. Interviews showed that students initially regarded discourse as a process of expressing and accepting ideas, rather than one involving build-on, challenging, and reviewing. However, after the program, student ZLY had an awareness of the importance of reflection and synthesizing.

|                    |  |
|--------------------|--|
| Pretest Interview  | "A good discourse is expressing ideas and accepting others' ideas."  |
| Posttest Interview | "A good discussion requires everyone to <i>reflect and synthesize these ideas in order to see if there any new problems emerged</i> ... We have different ideas in a discussion, and it is good to <i>generate new questions for further discussion</i> ..." |

#### From viewing good discourse as sharing of ideas to advancing community knowledge

To create knowledge, students not only need to understand the sharing of ideas, but also to distinguish the difference between knowledge sharing and knowledge creation. Students initially thought that discussion was for ideas sharing and problem solving only, and had no awareness of creating knowledge. Later on, when students engaged in the classroom meta-talk on principles, they started to emphasize on the difference between "sharing" and "creation". This excerpt shows that student GLM began to distinguish the two metaphors.

|                    |  |
|--------------------|--|
| Pretest Interview  | "Everyone share ideas...I am not sure whether we can create new knowledge..."  |
| Posttest Interview | "A good discussion needs to be a <i>continuous build-on</i> . We discussed " <i>knowledge contributor</i> " and " <i>knowledge creator</i> "... <i>for building on to advance community knowledge...and create new knowledge through discussion</i> ..." |

Students' interview responses were coded into three levels. Quantitative analyses indicated that students' understanding of discourse by interview correlated with their productive KF discourse moves, in terms of sustained inquiry ( $r=.556, p<.01$ ), enriching an explanation ( $r=.722, p<.01$ ), supporting an explanation ( $r=.439, p<.05$ ), improving an explanation ( $r=.738, p<.01$ ), challenging ( $r=.485, p<.05$ ), connection ( $r=.633, p<.01$ ), rise-above ( $r=.467, p<.05$ ), and synthesizing ( $r=.754, p<.01$ ). The findings suggested that students with deeper understanding of the principles also seemed more likely to be engaged in KF productive discourse.

## Conclusion and implications

The study has examined how students' collective inquiry of KB principles for developing productive discourse can be scaffold and its relationship to productive KF discourse. Analysis of KF discourse using KBDeX indicated how students' discourse became more coherent and productive in Period 2, after they inquiring on the KB principles in the classroom meta-talk, reflecting on their KF discourse and linking it to KB principles. Classroom design and processes suggested how students can engage in collective inquiry into KB principles to support their knowledge-building work. We also characterized students' understanding of principles as it related to their productive KF discourse engagement and examined relations between their understanding of principles with KF collective inquiry. Many studies have proposed the importance of KB principles, however, no systematic work were conducted on students' collective inquiry into principles using both KF and classroom discourse. This study has shown how students authentically worked on principles that were explicitly integrated with their KF discourse, and the possibility of changing students' understanding of the nature of discourse by linking it with principles. In sum, this study is particularly important, as little research has focused on how students' collective inquiry of KB principles can be scaffold to help them engage in productive discourse.

## References

- Bereiter, C., & Scardamalia, M. (2016). "Good Moves" in knowledge-creating dialogue. *QWERTY, 11*(2), 12-26.
- Chan, C. K. K., Lam, I. C. K., & Leung, R. W. H. (2012). Can collaborative knowledge building promote both scientific processes and science achievement? *International Journal of Educational Psychology, 1*(3), 199-227.
- Chuy, M., Zhang, J., Resendes, M., Scardamalia, M., & Bereiter, C. (2011). Does contributing to a knowledge building dialogue lead to individual advancement of knowledge. Paper presented at the CSCL2011.
- Hakkarainen, K. (2003). Progressive inquiry in a computer-supported biology class. *Journal of Research in Science Teaching, 40*(10), 1072-1088.
- Mercer, N., & Littleton, K. (2007). *Dialogue and development of children's thinking: A socio-cultural approach*. London: Routledge.
- Niu, H., & van Aalst, J. (2009). Participation in knowledge-building discourse: an analysis of online discussions in mainstream and honours social studies. *Canadian Journal of Learning and Technology, 35*(1), 1-23.
- Oshima, J., Oshima, R., & Matsuzawa, Y. (2012). Knowledge Building Discourse Explorer: a social network analysis application for knowledge building discourse. *Educational Technology Research and Development, 60*(5), 903-921.
- Sawyer, R. K. (2014). *The Cambridge handbook of the learning sciences*. New York, NY: Cambridge University Press.
- Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Ed.), *Liberal education in a knowledge society* (pp. 67-98). Chicago: Open Court.
- Scardamalia, M., & Bereiter, C. (2014). Knowledge building and knowledge creation: Theory, pedagogy, and technology. In R. K. Sawyer (Eds.), *The Cambridge handbook of the learning sciences* (2<sup>nd</sup> ed., pp. 397-417). New York, NY: Cambridge University Press.
- Stahl, G., Cress, U., Ludvigsen, S., & Law, N. (2014). Dialogic foundations of CSCL. *International Journal of Computer-Supported Collaborative Learning, 9*(2), 117-125.
- van Aalst, J., & Chan, C. K. K. (2007). Student-directed assessment of knowledge building using electronic portfolios. *The Journal of the Learning Sciences, 16*(2), 175-220.
- Wegerif, R. (2001). Applying a dialogical model of reason in the classroom. In Joiner, R., Faulkner, D., Miell, D., & Littleton, K. (Eds.) *Rethinking Collaborative Learning*, Free association Press.
- Zhang, J., Scardamalia, M., Lamon, M., Messina, R., & Reeve, R. (2007). Socio-cognitive dynamics of knowledge building in the work of 9- and 10-year-olds. *Educational Technology Research & Development, 55*, 117-145. *Journal of the Learning Sciences, 27*(3), 390-430.
- Zhang, J., Tao, D., Chen, M.-H., Sun, Y., Judson, D., & Naqvi, S (2018). Co-organizing the collective journey of inquiry with idea thread mapper. *Journal of the Learning Sciences, 27*(3), 390-430.



# Indexes



# Author Index

Pages 1-542: Volume 1  
Pages 543-1104: Volume 2

- Abassi, Nour Ghalia, 783  
Adeoye, Temitope, 775, 867  
Agesilaou, Andria, 751  
Agrawal, Harshit, 488  
Ahmed, Ishrat, 545  
Ahn, June, 605, 1041  
Alberto, Rosa, 660  
Aldemir, Tugce, 320  
Aleman, Ezequiel, 905  
Aleven, Vincent, 977  
Alexandre, Jeremy, 9  
Allert, Heidrun, 168  
Alozie, Nonye, 628  
Alqahtani, Mona, 589  
Amador, Julie, 569  
Amundrud, Anja, 648  
An, Haokang, 855  
Anderson, Emma, 821, 849, 889  
Andrade, Alejandro, 955  
Andrews, Christopher, 33, 656, 775  
Andrews-Todd, Jessica, 565, 1035  
Ang, JL, 561  
Angel, Dror, 735  
Anthony, Lisa, 9  
Aridor, Keren, 735  
Aritajati, Chulakorn, 296  
Ashley, Robert, 688  
Atias, Osnat, 735, 843  
Axisa, Carmen, 859  
Baabdullah, Afaf, 853  
Bae, Haesol, 947, 949, 1025  
Bailie, Addisu, 941  
Baker, Michael, 601  
Bakker, Arthur, 660  
Baram-Tsabari, Ayelet, 735, 843  
Barany, Amanda, 216, 1036  
Barbaro, Valerie, 857  
Barksdale, Sarah, 616, 825  
Barron, Brigid, 344  
Bates, Meg, 577  
Bayat, Arash, 809  
Beilstein, Shereen, 577  
Beirl, Diana, 96  
Belcher, John, 997  
Beldame, Magali Ollagnier, 978  
Bell, Adam, 799, 1082  
Bell, Amanda, 336, 775  
Bellhäuser, Henrik, 719  
Benachour, Phil, 925  
Ben-Horin, Hava, 735  
Benichou, Maya, 735, 843  
Ben-Zvi, Dani, 735  
Bereiter, Carl, 767  
Bers, Marina, 652  
Bichler, Sarah, 668  
Bielaczyc, Katerine, 1079  
Biemans, Harm, 644  
Bietti, Lucas, 601  
Biswas, Gautam, 360  
Blikstein, Paulo, 1021  
Bodemer, Daniel, 25, 719, 1094  
Bogart, Christopher, 719  
Bonsignore, Elizabeth, 605, 1041  
Boomgaarden, Antje, 863  
Borar, Priyanka, 488  
Borge, Marcela, 320, 589, 620, 945, 975, 1043  
Bottineau, Didier, 978  
Boulden, Danielle, 224, 833  
Boyer, Kristy Elizabeth, 224, 833  
Bracey, Zoë Buck, 743  
Bradford, Allison, 480  
Brady, Corey, 512, 704, 775, 879, 901  
Bratt, Harry, 628  
Bressler, Denise, 821  
Bridges, Susan, 915  
Briskin, Jessica, 700  
Broisin, Julien, 977  
Bryant, Tonya, 923  
Buckingham Shum, Simon, 859  
Burriss, Sarah, 835  
Bussian, Christine, 168  
Cao, Shengyang, 376  
Carrillo, Rubiella, 861  
Carson, Cynthia, 569  
Cassidy, Robert, 448  
Castek, Jill, 939  
Cavera, Veronica, 73  
Ceja-Salgado, Omar, 917  
Cervantes, Alexandria, 336  
Chai, Aileen S.W., 561  
Chai, Ching Sing, 767, 815  
Chaivisit, Yam, 911  
Chan, Carol K.K., 288, 534, 767, 959, 1045, 1047, 1048, 1092  
Chan, Kennedy, 959  
Chan, Lap Ki, 915  
Chandrasekharan, Sanjay, 488  
Chandy, Betty, 827



Chang, Peter, 9  
 Chang, Yu-Hui, 616, 791, 839  
 Chapman, Katherine, 791, 1009  
 Chapman, Susan, 967  
 Charles, Elizabeth, 448  
 Chartrand, Grant, 33, 656, 775  
 Chen, Gaowei, 959, 1092  
 Chen, Jennifer King, 708  
 Chen, Laure Lu, 416  
 Chen, Mei-Hwa, 144, 456, 1045  
 Chen, Wenli, 240, 561, 977  
 Chen, Xiangdong, 907  
 Chen, Yuxin, 893, 1029  
 Cheng, Bosco, 392  
 Cheng, Britte, 775, 867  
 Cheng, Britte Haugan, 1040  
 Ching, Cynthia Carter, 799  
 Choppin, Jeffrey, 569  
 Chounta, Irene-Angelica, 1066  
 Christidou, Dimitra, 759  
 Chu, Sharon Lynn, 1074  
 Cimpian, Joseph, 577  
 Claire, Polo, 424  
 Clegg, Tamara, 605, 1041  
 Coiro, Julie, 939  
 Conlin, Luke, 360  
 Constantine, Angelina, 616  
 Corning, Wiley, 997  
 Craig, Kalani, 949, 1025  
 Crossley, Scott, 640  
 Dabholkar, Sugat, 827, 845  
 Dai, Zhaihuan, 696  
 Dai, Zhaohui, 899  
 Damsa, Crina, 775, 1037, 1043  
 D'Angelo, Cynthia, 628  
 Danish, Joshua, 408, 676, 799, 949, 953, 1025, 1079  
 Darrow, Alice, 9  
 Dascalu, Mihai, 640, 877  
 Dasgupta, Chandan, 664  
 Dashti, Cameron, 719, 855  
 Davidsen, Jacob, 17, 759  
 Davis, Bria, 676, 799  
 Davis, Katie, 176  
 Davis, Pryce, 905, 1043  
 de Jong, Ton, 432  
 Deng, Xiaozhou, 232  
 Dillenbourg, Pierre, 977, 1013  
 Dimitriadis, Yannis, 813, 1005  
 DiSalvo, Elizabeth, 1041  
 Dobbie, Karen, 81  
 Doleck, Tenzin, 861  
 Dolonen, Jan Arild, 965  
 Donaldson, Jonan Phillip, 897, 1036  
 Doorman, Michiel, 660  
 Dresel, Markus, 312  
 Drijvers, Paul, 660  
 Dubovi, Ilana, 88, 593, 799  
 Dugdale, Michael, 448  
 Duncan, Ravit Golan, 73  
 Eagan, Brendan, 136, 328, 608, 1042  
 Easterday, Matthew, 811, 819  
 Eberle, Julia, 1043, 1051  
 Echeverria, Vanessa, 859  
 Ehrenfeld, Nadav, 120  
 Eie, Siv, 272  
 Elliott, Doug, 859  
 Emara, Mona, 360  
 Englander, Claire, 929  
 Englund, Jennifer, 616  
 Enyedy, Noel, 676, 799, 953  
 Er, Erkan, 813, 1005, 1070  
 Erkan, Betul, 865  
 Erkens, Melanie, 719  
 Eshuis, Elise, 719  
 Espino, Danielle, 608, 612  
 Evans, Abigail, 176  
 Eysink, Tessa, 719  
 Fair, Jordan, 929  
 Farrell, Cayley, 328  
 Faucon, Louis, 1013  
 Feng, Xueqi, 288  
 Fields, Deborah, 472, 799  
 Filliettaz, Laurent, 969  
 Finlay, Michael, 817  
 Fiore, Stephen, 3  
 Fischer, Frank, 668, 727, 783, 1043  
 Fischer, Martin, 727  
 Forkner, Michael, 520  
 Fortune, Angela, 751  
 Foster, Aroutis, 216  
 Fu, Xiaoting, 833  
 Gadille, Martine, 549  
 Gasevic, Dragan, 813, 1005  
 Gebre, Engida, 941  
 Georgen, Chris, 280, 676, 955  
 Georgiou, Yiannis, 304, 1068  
 Gerard, Libby, 743, 885  
 Gerber, Elizabeth, 811, 819  
 Giannakos, Michail, 684  
 Gijlers, Hannie, 719  
 Gillespie, Ryan, 569  
 Gillet, Denis, 432  
 Giri, Sagun, 688  
 Girotto, Victor, 545  
 Glazewski, Krista, 893, 947, 1029  
 Gleaves, Schuyler, 9  
 Gnesdilow, Dana, 520, 672  
 Gnoli, Alessandro, 981  
 Goggins, Sean, 1017  
 Golan, Oren, 735  
 Goldman, Susan R., 751, 1048  
 Goldstone, Robert, 1039  
 Golumbic, Yaela N., 735

Gomez, Kimberley, 751, 1048  
 Gomez, Louis, 751  
 Gomoll, Andrea, 751, 775, 867, 903, 1084  
 Gonzalez, Carmen, 160  
 Graesser, Art, 727  
 Grauwin, Sebastian, 712, 935  
 Graville, Cynthia, 929  
 Green, Judith, 915  
 Greenwald, Scott W., 997  
 Greiff, Samuel, 727  
 Greisel, Martin, 312  
 Gresalfi, Melissa, 336, 775  
 Griffin, Joseph, 1017  
 Grover, Shuchi, 360  
 Gu, Xiaoqing, 971  
 Guillain, Leonore Valentine, 783  
 Guo, Chen, 907  
 Gutu-Robu, Gabriel, 877  
 Håklev, Stian, 783, 1013, 1076  
 Hall, Allison H., 751  
 Hall, Rogers, 759, 799  
 Halverson, Erica, 899  
 Hamilton, Eric, 608, 612, 1042  
 Hao, Jiangang, 727, 1035  
 Harrison, Emily, 885, 927  
 Hasenbein, Lisa, 668  
 Hatami, Javad, 644, 809  
 Haudek, Kevin C. , 743  
 Hays, Maria, 799  
 Hei, Cheuk, 392  
 Heimbuch, Sven, 25  
 Henricks, Genevieve, 577  
 Henry, Jessica, 636  
 Hernández-Leo, Davinia, 248, 891  
 Hestness, Emily, 869  
 Hickey, Daniel T., 33, 656, 775  
 Higgs, Jennifer, 1058  
 Hilton, Michael, 719, 855  
 Hmelo-Silver, Cindy, 712, 751, 775, 783, 867, 893, 895, 915, 947, 949, 1025, 1029  
 Hoadley, Christopher, 735  
 Hod, Yotam, 735, 975, 1079  
 Holland, Beth, 939  
 Hong, Huang-Yao, 632, 692, 767, 815  
 Hoppe, H. Ulrich, 719  
 Horn, Ilana, 120  
 Housh, Karyn, 893  
 Hu, Liru, 1092  
 Huang, Joey, 933  
 Hulsbergen, Michiel, 877  
 Humburg, Megan, 676, 799  
 Hutchins, Nicole, 360  
 Ikubo, Shinya, 873  
 Ikeda, Shono, 851  
 Impedovo, Maria Antonietta, 549  
 Ioannou, Andri, 192, 304, 384, 978  
 Ioannou, Marianna, 304  
 Itow, Rebecca, 775  
 Jacque, Berri, 636  
 Jain, Rupal, 823  
 Jasien, Lara, 791  
 Jay, Victoria, 577  
 Jayathirtha, Gayithri, 472, 799  
 Jensen, Bryant, 557  
 Jeong, Heisawn, 712, 783, 935, 1079  
 Jeong, Jiyeon, 712  
 Jeurung, Johan, 877  
 Jiang, Shiyan, 208, 1054  
 Jin, Qiao, 232  
 Johnson, Mina, 304  
 Johnstun, Kevin, 557  
 Kafai, Yasmin, 440, 472, 526, 799  
 Kahn, Jennifer, 208  
 Kajamaa, Anu, 352, 585  
 Kali, Yael, 735, 843, 975, 1037  
 Kanasa, Harry, 967  
 Karnam, Durgaprasad, 488  
 Karumbaiah, Shamyra, 827  
 Kaur, Navneet, 664  
 Kazemitabar, Maedeh, 861  
 Ke, Fengfeng, 696  
 Keifert, Danielle, 953  
 Kent, Carmel, 735  
 Ketelhut, Diane Jass, 869  
 Ketenci, Tuba, 865  
 Keune, Anna, 913, 961, 1039  
 Khanlari, Ahmad, 767, 963  
 Kharrufa, Ahmed, 1044  
 Kidron, Adi, 735, 1037  
 Kiili, Carita, 939  
 Kim, Jee-Seon, 520  
 Kim, Joomi, 935  
 Kim, Min Kyu, 883  
 Kirkland, Brittani, 9  
 Ko, Monlin Monica, 751  
 Koedinger, Ken, 504  
 Koh, Elizabeth, 1045  
 Kollar, Ingo, 312  
 Kolodner, Janet L. , 1048  
 Kolovou, Maria, 909  
 Konert, Johannes, 719  
 Kopcha, Theodore, 464  
 Koponen, Ismo, 837  
 Korhonen, Jasmiina, 1090  
 Krämer, Nicole, 256  
 Kumar, Vishesh, 680  
 Kumpulainen, Kristiina, 352, 585  
 Kyza, Eleni, 751  
 Lachand, Valentin, 989  
 Laffey, James, 1017  
 Lagrange-Lanaspre, Sandra, 424  
 Lajoie, Susanne, 861  
 Lala, Raja, 877  
 Land, Susan, 700

Lander, Sean, 1017  
 Lanouette, Kathryn, 553  
 Lapham, Sam, 392  
 Latifi, Saeed, 644  
 Lavoué, Elise, 296, 861  
 Law, Nancy, 416, 1051  
 Lawrence, LuEttaMae, 829  
 Leary, Heather, 557  
 Lechelt, Zuzanna, 875  
 Lee, Changhee, 939  
 Lee, Hee-Sun, 743  
 Lee, Hyeyeon, 945  
 Lee, Jiyeon, 767  
 Lee, Seung, 608, 612  
 Lee, Seung Bok, 1042  
 Lee, Victor R., 88, 799, 1051  
 Lenton, Kevin, 448  
 Lester, James, 893, 947, 1029  
 Leuders, Timo, 863  
 Lewis, Daniel Rees, 811, 819, 1060  
 Ley, Tobias, 432  
 Li, Haiying, 727  
 Lian, Zheming, 112  
 Lin, Feng, 597  
 Lin, Pei-Yi, 632, 692, 767  
 Lin, Veronica, 1021  
 Lindberg, Lindsay, 799, 953  
 Lindgren, Robb, 799  
 Linn, Marcia C., 49, 480, 668, 708, 743, 885, 927, 1048  
 Liru, Hu, 959  
 Litts, Breanne, 799  
 Liu, Ou Lydia, 743  
 Loibl, Katharina, 863  
 Lommi, Henri, 837  
 Looi, Chee-Kit, 240, 561, 841  
 Loparev, Anna, 652  
 Lopez Silva, Brenda, 1038  
 Lott, Kurt, 929  
 Luc, Annie, 9  
 Ludvigsen, Sten Runar, 775, 965, 975, 1047  
 Lui, Debora, 799  
 Lui, Michelle, 1038  
 Lund, Kristine, 935  
 Lundh, Patrik, 775, 867  
 Lynch, Collin, 224, 833  
 Ma, Leanne, 81, 632, 881, 943  
 Ma, Wanjing (Any), 49  
 Maes, Pattie, 997  
 Manathunga, Kalpani, 248  
 Manor, Hana, 735  
 Manske, Sven, 719  
 Markaki-Lothe, Vasiliki, 969  
 Markauskaite, Lina, 1037  
 Marquardt, Nicolai, 875  
 Martin, Caitlin, 344  
 Martin, Kit, 931  
 Martin, Nicole, 672  
 Martinez-Maldonado, Roberto, 859, 1044  
 Matsuzawa, Yoshiaki, 881  
 Matuk, Camillia, 49  
 Mawasi, Areej, 545  
 McBride, Elizabeth, 927, 1062  
 McCranie, Ann, 1025  
 McDonough, Emily Kate, 636  
 McDowell, Gavin, 997  
 McElroy-Brown, Kiley, 73  
 McGartland, Tessa, 929  
 McNamara, Danielle, 640, 743, 977  
 Melzner, Nadine, 312  
 Mercier, Emma, 581, 829, 871, 975  
 Merrill, BriYana, 929  
 Michel, Christine, 989  
 Mifsud, Louise, 272  
 Miller, Brian W., 817  
 Miller, Katherine, 821, 849, 889  
 Mills, Kelly, 605, 869  
 Mishol-Shauli, Nakhi, 735  
 Mishra, Shitanshu, 1072  
 Mitros, Piotr, 1035  
 Moher, Tom, 981, 1038  
 Molinari, Gaëlle, 977  
 Mondada, Lorenza, 4  
 Moore, Darby, 929  
 Moore, Meredith, 823  
 Moore, Monika, 200  
 Morales, Amanda, 9  
 Moran, Cheryl, 577  
 Mørch, Anders, 272  
 Morgan, Taylor, 929  
 Morganfield, Kennedy, 929  
 Morphew, Jason, 799  
 Mortensen, Chase, 799  
 Mott, Bradford W., 947, 1029  
 Mueller, Amy, 899  
 Müller, Adrienne, 719  
 Muukkonen, Hanni, 1037  
 Naglieri, Raymond, 696  
 Nathan, Mitchell J., 1047  
 Neff, Hannah, 9  
 Nerland, Monika, 1037  
 Ngoh, Clarence, 855  
 Nguyen, Judy, 344, 496  
 Nichols-Paez, Isaac, 879  
 Noroozi, Omid, 644, 809  
 Noushad, Noora, 937  
 O'Brien, Annie, 929  
 Ocak, Ceren, 464  
 Ogan, Amy, 504  
 Ohsaki, Ayano, 57  
 Ollesch, Lisa, 1094  
 Olsen, Jennifer, 400, 977, 1013  
 Ong, Aloysius, 1045  
 Osborne, Jonathan, 743

Oshima, Jun, 57, 851  
 Oshima, Ritsuko, 57, 851  
 Ouyang, Fan, 831, 839, 1064  
 Oztok, Murat, 925  
 Pachman, Mariya, 696  
 Pallant, Amy, 743  
 Palonen, Tuire, 775  
 Papadopoulos, Pantelis M., 727, 1001  
 Papavlasopoulou, Sofia, 684  
 Pappas, Ilias, 684  
 Park, Hyejin, 624  
 Park, Seung Won, 416  
 Patel, Aastha, 664  
 Pauw, Daniel, 605  
 Peppler, Kylie, 408, 913, 933, 961, 1039  
 Perry, Judy, 957  
 Perry, Michelle, 577  
 Pesaresi, Simona, 573  
 Peterson, Lana, 616, 825  
 Petrosino, Tony, 901  
 Phuong, Andrew, 496  
 Pierroux, Palmyre, 759  
 Pierson, Ashlyn, 1086  
 Pina, Laura, 160  
 Pinkard, Nichole, 1048  
 Pion, Carmit, 735  
 Pitt, Caroline, 605  
 Planey, James, 799  
 Plummer, Julia, 887  
 Pollalis, Christina, 652  
 Polman, Joseph, 929  
 Polo, Claire, 978  
 Power, Tamara, 859  
 Pozen, Rebecca, 136  
 Praharaaj, Sambit, 1045  
 Pressman, Emily, 751  
 Price, Sara, 917  
 Prieto, Luis P. , 432  
 Proctor, Chris, 104  
 Puntambekar, Sadhana, 520, 597, 672  
 Radinsky, Joshua, 200, 1047  
 Radkowsch, Anika, 727, 783  
 Radu, Iulian, 128, 264, 923  
 Rafaeli, Sheizaf, 735  
 Raffel, Lars-Arne, 168  
 Ramey, Kay, 368  
 Rasmussen, Ingvill, 648, 965, 993  
 Rattenni, Alize, 392  
 Rechavi, Amit, 735  
 Recker, Mimi, 1079  
 Rehmat, Abeera, 893  
 Reichelt, Norma, 168  
 Reichsman, Frieda, 73  
 Reider, David, 821, 849, 889  
 Reimann, Peter, 1037  
 Reimer, Paul, 791  
 Reimers, Jackson, 704  
 Resendes, Monica, 81, 943  
 Resnick, Lauren, 959  
 Reynolds, Caroline, 392  
 Richard, Gabriela, 688  
 Richards, Sonya, 668  
 Richey, Colleen, 628  
 Richter, Christoph, 168  
 Riesland, Erin, 799  
 Riyami, Bouchaib, 847  
 Roberto, Collette, 951  
 Rodela, Katherine, 751  
 Rodríguez-Triana, María Jesús, 432  
 Rogat, Toni Kempler, 775, 867, 1040  
 Rogers, Yvonne, 96, 875  
 Roldan, Wendy, 160  
 Röpke, René, 719  
 Roschelle, Jeremy, 1047  
 Rosé, Carolyn Penstein, 719, 855, 977  
 Rosson, Mary Beth, 296  
 Rowe, Jonathan, 947, 1029  
 Rowsell, Jennifer, 1039  
 Roy, Jean-Michel, 978  
 Ruis, Andrew, 112, 136, 376  
 Rummel, Nikol, 975, 977, 1048  
 Ruseti, Stefan, 877  
 Rush, Adam, 929  
 Ryberg, Thomas, 17  
 Sadler, Troy, 1017  
 Sagy, Ornit, 843, 975  
 Saito, Moegi, 873  
 Sakr, Majd, 719, 855  
 Saleh, Asmalina, 893, 915, 1029, 1052  
 Sankaranarayanan, Sreecharan, 719, 855  
 Sanzenbacher, Beth, 1038  
 Scardamalia, Marlene, 81, 767, 881, 963, 1045  
 Scharber, Cassie, 616, 825  
 Schejter, Amit, 735  
 Schmidmaier, Ralf, 727  
 Schmitt, Lara, 719  
 Schneider, Bertrand, 41, 128, 264, 923, 1044  
 Schneider, Leslie, 636  
 Schoenfeld, Ilana, 821, 849, 889  
 Schuman, Carrie, 9  
 Schwarz, Baruch, 975  
 Sedas, R. Mishael, 961  
 Seet, Chih Hui, 815  
 Semerjian, Amy, 951  
 Seo, Sujong, 712  
 Shacham, Malka, 735  
 Shaer, Orit, 652  
 Shaffer, David Williamson, 112, 136, 152, 328, 376, 1037, 1102  
 Shah, Mamta, 216  
 Shanahan, Patrick, 929  
 Shapiro, Ben Rydal, 184, 759, 1009, 1056  
 Sharma, Garima, 49  
 Sharma, Kshitij, 400, 684, 977

Shaw, Mia, 526, 799  
 Shehab, Saadeddine, 581, 871  
 Sheldon, Josh, 985  
 Shen, Ji, 909  
 Sherard, Maximilan, 901  
 Sherin, Bruce, 743  
 Shim, Jooeun, 821, 827, 849, 889, 937  
 Shirouzu, Hajime, 873  
 Shvarts, Anna, 660  
 Siebert-Evenstone, Amanda, 136, 152, 1042  
 Sigoloff, Justin, 1017  
 Silvis, Deborah, 65, 799  
 Sirbu, Maria-Dorinela, 640  
 Sivaraj, Ramya, 616  
 Slater, Stefan, 440  
 Slotta, James, 448, 636, 1038  
 Smith, Blaine, 909  
 Smith, Jennifer, 628  
 Smørdal, Ole, 993  
 Snyder, Caitlin, 360  
 Socratous, Chrysanthos, 192  
 Sommerhoff, Daniel, 1043  
 Soni, Nikita, 9  
 Sparks, Jesse R. , 939  
 Splichal, Jin, 57  
 Stadler, Matthias, 727  
 Steier, Rolf, 759  
 Stevens, Reed, 368  
 Stofer, Kathryn A. , 9  
 Stoyanova, Filipa, 256  
 Strauß, Sebastian, 1098  
 Strawhacker, Amanda, 652  
 Stucker, Michael, 955  
 Sullivan, Amanda, 652  
 Sun, Jingjing, 817  
 Sun, Na, 296  
 Sunbury, Susan, 887  
 Sung, Hanall, 376  
 Swenson, Erik, 929  
 Swiecki, Zachari, 112, 328, 1042, 1102  
 Sylar, Ben, 929  
 Szostalo, Maksymilian, 949, 1025  
 Tabak, Iris, 200, 593, 735, 751  
 Tabard, Aurélien, 296, 989  
 Talafian, Hamideh, 216  
 Tan, Esther, 1045  
 Tan, Seng Chee, 767  
 Taylor, Katie Headrick, 799  
 Taylor, Robert, 1029  
 Tchounikine, Pierre, 975  
 Teasedale, Michael, 929  
 Teasley, Stephanie D., 977, 1048  
 Teo, Chew Lee, 1045  
 Thompson, Kate, 967, 1037  
 Thompson, Naomi, 408, 1096  
 Thompson, Penny, 911  
 Timotheou, Stella, 384  
 Tirosh, Noam, 735  
 Tissenbaum, Catherine Dornfeld, 520, 672  
 Tissenbaum, Mike, 680, 985  
 Tong, Yuyao, 534, 767  
 Toprani, Dhvani, 589  
 Toulis, Andrew, 943  
 Trausan-Matu, Stefan, 640, 1036  
 Traynor, Anne, 775, 867  
 Tsai, Chin-Chung, 767  
 Tsan, Jennifer, 224, 833  
 Tsang, Haley, 416  
 Tsiartas, Andreas, 628  
 Tsovaltzi, Dimitra, 719  
 Tu, Xintian, 676, 799  
 Tubman, Philip, 925  
 Turkkila, Miiikka, 837  
 Uchidiuno, Judith, 504  
 Udomprasert, Patricia, 887  
 Unahalekhaka, Apittha, 264  
 Uttamchandani, Suraj, 895, 949, 961, 1025, 1100  
 Vaishampayan, Abha, 887  
 van Aalst, Jan, 288, 767, 1047  
 van Dortmund, Jordy, 877  
 van Geest, Marcell, 877  
 Van Tress, Lauren, 612  
 Van Wart, Sarah, 553  
 Vandenberg, Jessica, 224, 833  
 Vanegas, Paola, 160  
 Vargas-Vite, Emilio, 504  
 Venant, Rémi, 977  
 Verish, Clarissa, 652  
 Vidal, Phillipe, 977  
 Vogel, Freydis, 783, 1043  
 Vogelstein, Lauren, 512, 775, 791  
 Vujovic, Milica, 891  
 Walker, Erin, 545  
 Walker, Justice, 440, 526  
 Walker, Susan, 392  
 Wan, Sally Wai Yan, 783, 921  
 Wan, Sancia Wai-San, 783  
 Wang, Chunli, 971  
 Wang, Danli, 232  
 Wang, Xu, 719, 855  
 Wang, Yinying, 883  
 Weinberger, Armin, 719  
 Weiss, Patrice L. , 735  
 Wen, Yun, 240, 561, 841  
 Wendel, Daniel, 821, 849, 889  
 Westendorf, Lauren, 652  
 Wheeler, Penny, 1036  
 White, Erikah, 929  
 Whitehurst, Amanda, 545  
 Whittington, Nakisha, 688  
 Wiebe, Eric, 224, 833  
 Wilensky, Uri, 845  
 Wiles, Janet, 5  
 Wiley, Korah, 480

Wilkinson, Jamiaka, 833  
Wilson, Chris, 743  
Wise, Alyssa, 975, 977  
Wobbrock, Jacob, 176  
Wohlwend, Karen, 1039  
Womack, Andrew, 1017  
Worsley, Marcelo, 1044  
Wu, Jiajun, 959  
Wulff, Eric, 1017  
Wylie, Ruth, 545  
Xia, Yu, 320, 620, 945  
Xu, Xinhao, 696  
Yang, Yuqin, 288  
Yarzebinski, Evelyn, 504  
Yavuz, Sinan, 520  
Ye, Andy, 827  
Yeo, Jennifer, 240  
Yip, Jason C., 160, 605, 1041  
Yoon, Susan, 783, 821, 827, 849, 889, 937, 1047  
Yu, Jinjian, 959  
Yuan, Guangji, 456, 1088  
Yuan, Xin, 696  
Yuill, Nicola, 96  
Zakaria, Zarifa, 224  
Zapata-Rivera, Diego, 1035  
Zhang, Chao, 448  
Zhang, Jianwei, 144, 456, 573, 624, 767, 1045,  
1088  
Zhang, Si, 240, 561  
Zhang, Yibing, 767  
Zhu, Gaoxia, 943  
Zhu, Mengxiao, 565, 1035  
Zuckerman, Oren, 735

# Keyword Index

Pages 1-542: Volume 1  
Pages 543-1104: Volume 2

|  |  |
|--|--|
| access.....                                    | 895  |
| action research.....                           | 549, 865, 873, 925, 929  |
| adoption of technologies.....                  | 96, 304, 384, 955  |
| affective dimensions.....                      | 304, 861, 951  |
| argumentation.....                             | 240, 526, 644, 817   |
| artifact analysis.....                         | 743, 909   |
| artificial intelligence.....                   | 640, 877, 955  |
| augmented/virtual reality.....                 | 128, 264, 280, 676, 841, 917, 931, 953, 957  |
| best practices.....                            | 384, 448   |
| broadening participation.....                  | 168, 368, 616  |
| case studies.....                              | 208, 216, 320, 336, 384, 448, 464, 480, 605, 616, 636, 751, 791, 815, 817, 825, 901, 903, 921, 937, 945, 957   |
| cognitive dimensions.....                      | 464, 520, 767, 799   |
| cognitive psychology.....                      | 216, 416   |
| cognitive studies.....                         | 464, 520   |
| collaboration skills.....                      | 96, 440, 589, 608, 612, 644, 656, 672, 813, 831, 853, 861, 867, 883, 891, 895, 939, 971  |
| collaborative problem solving skills.....      | 41, 112, 128, 160, 192, 224, 232, 312, 360, 376, 400, 416, 561, 612, 616, 652, 664, 819, 893, 923, 933   |
| collective regulation/socio-metacognition..... | 57, 192, 573, 775, 813, 851, 945   |
| communication studies.....                     | 877  |
| community building.....                        | 392, 577, 608, 831, 849, 889   |
| community settings.....                        | 392, 735   |
| computational models.....                      | 360, 512, 676, 704, 845  |
| computer science.....                          | 248, 472, 825, 933   |
| computing education.....                       | 104, 224, 232, 320, 336, 464, 512, 616, 680, 688, 815, 825, 827, 833, 855, 865, 901, 913, 933, 951   |
| connected learning.....                        | 440  |
| constructivism/constructionism.....            | 192, 480, 831, 839, 879, 897, 961  |
| content analysis.....                          | 25, 144, 240, 296, 320, 561, 712, 825, 831, 857, 873, 909, 935, 945, 971   |
| contextual design.....                         | 336, 887   |
| conversational analysis.....                   | 612, 969   |
| creativity.....                                | 456, 549, 585  |
| cross-cultural studies.....                    | 504, 688, 712  |
| cross-sectional studies.....                   | 712  |
| Cultural-Historical Activity Theory.....       | 799  |
| culturally relevant pedagogy/design.....       | 104, 504, 601, 817   |
| culture.....                                   | 897, 921   |
| descriptive studies.....                       | 96, 520, 865, 887  |
| design experiments.....                        | 73, 88, 644, 873, 925  |
| design thinking.....                           | 408, 664, 692, 811, 819, 903   |
| design-based research.....                     | 144, 176, 272, 456, 526, 553, 557, 636, 688, 704, 811, 815, 823, 845, 865, 869, 879, 905, 909, 925, 939, 967   |
| dialogism.....                                 | 965  |
| discourse analysis.....                        | 112, 136, 144, 224, 328, 376, 416, 565, 573, 624, 672, 833, 915, 965   |
| distance learning.....                         | 809  |
| distributed cognition.....                     | 472, 601, 704  |
| diversity and inclusion.....                   | 344, 496, 557, 605   |
| early childhood education.....                 | 504  |
| education technology.....                      | 168, 208, 216, 240, 304, 320, 384, 400, 456, 464, 585, 589, 636, 648, 684, 696, 708, 719, 727, 783, 809, 813, 829, 831, 845, 847, 849, 883, 891, 901, 917, 921, 927, 965 |

|   |  |
|---|--|
| educational psychology .....                    | 520, 597, 644, 668, 809  |
| engagement.....                                 | 352, 839, 883, 915   |
| engagement and transfer.....                    | 775, 867   |
| equity and gender.....                          | 368, 616, 905, 931   |
| ethnography.....                                | 216, 488   |
| ethnomethodology .....                          | 17   |
| exergaming .....                                | 304  |
| experimental studies .....                      | 320, 400, 644, 684, 696, 847   |
| game-based media .....                          | 73   |
| gaming .....                                    | 440, 652, 877, 893   |
| gesture and gaze.....                           | 17, 41, 424, 660, 923  |
| group cognition.....                            | 416, 424   |
| handheld/mobile devices .....                   | 799, 957   |
| history .....                                   | 272, 931, 949  |
| human-computer interaction.....                 | 25, 96, 296, 545, 561, 589, 859, 875   |
| identity and community .....                    | 549, 835   |
| informal learning settings .....                | 9, 65, 160, 208, 280, 312, 34, 416, 593, 608, 735, 791, 899, 905, 929, 931                               |
| information analysis .....                      | 743  |
| information sciences .....                      | 160  |
| information sharing.....                        | 49   |
| innovative algorithms .....                     | 943  |
| innovative design approach .....                | 700, 955   |
| innovative interfaces.....                      | 911, 939   |
| innovative software/artifacts.....              | 73, 636, 855, 945, 949   |
| inquiry learning .....                          | 49, 200, 668, 708, 843, 921, 939   |
| instrumentation .....                           | 660  |
| intelligent systems .....                       | 176, 628   |
| interaction analysis .....                      | 17, 65, 120, 184, 336, 352, 368, 400, 424, 565, 585, 589, 640, 648,<br>684, 688, 719, 727, 759, 875, 915 |
| interaction tool.....                           | 33, 248, 911   |
| intersubjectivity .....                         | 272, 965   |
| knowledge building .....                        | 81, 192, 288, 432, 534, 545, 573, 585, 624, 632, 692, 767, 851, 853, 881, 907, 943, 963                  |
| large-scale studies.....                        | 488, 727, 813  |
| learning analytics/educational data mining..... | 144, 240, 328, 400, 565, 628, 719, 727, 813, 859, 881, 891,<br>943, 947, 955, 959                        |
| learning analytics and monitoring tools.....    | 41, 288, 628, 640, 829, 837, 851, 891  |
| learning management systems .....               | 847  |
| learning outcomes.....                          | 25, 304, 496, 612, 885, 951  |
| literacy .....                                  | 104, 835, 909, 939   |
| longitudinal studies.....                       | 312, 640   |
| mathematics education .....                     | 120, 384, 488, 660, 863, 879, 961  |
| MOOCs.....                                      | 256, 656, 821, 847, 849, 889   |
| motivation.....                                 | 847, 857, 951  |
| multi-touch devices.....                        | 9, 488   |
| online learning .....                           | 33, 256, 296, 376, 656, 700, 735, 827, 839, 849, 857   |
| participatory design .....                      | 57, 545, 549, 751, 827, 929  |
| participatory simulations .....                 | 512  |
| portable/ubiquitous collaboration .....         | 65, 911  |
| primary school.....                             | 352, 585, 632, 648, 676, 833, 875  |
| problem-based learning .....                    | 17, 120, 581, 819, 863, 871, 893, 903, 915, 947, 951, 967  |
| process support.....                            | 296, 480, 767, 895, 899, 965   |
| project-based instruction.....                  | 811, 855   |
| psychometric analysis.....                      | 727  |
| quantified self.....                            | 684, 799   |
| reading .....                                   | 408, 632, 907  |
| reflection.....                                 | 192, 823, 851, 859, 935, 959   |
| regulation.....                                 | 176  |



|  |   |
|--|---|
| representational affordances .....     | 553, 845, 961   |
| representational tools.....            | 200, 759, 799, 809, 927, 941  |
| research methodologies .....           | 57, 112, 184, 328, 608, 612, 620  |
| scaffolding .....                      | 88, 96, 200, 534, 597, 648, 783, 809, 863, 907, 915   |
| science education.....                 | 9, 49, 57, 73, 128, 152, 216, 240, 264, 280, 288, 360, 408, 456, 520, 526, 553, 561,<br>565, 597, 636, 652, 668, 672, 676, 708, 743, 821, 825, 843, 845, 873,<br>885, 887, 901, 909, 927, 937, 941, 947, 953, 955 |
| scripting .....                        | 248, 561  |
| secondary school.....                  | 472, 488, 534, 680, 937, 967  |
| self-regulated learning .....          | 891, 945  |
| semiotics .....                        | 887   |
| shared knowledge .....                 | 352, 432, 624, 767  |
| simulations.....                       | 88, 152, 696, 859, 885, 877   |
| situated learning.....                 | 33, 152, 696, 775, 791, 867, 969  |
| skill development.....                 | 545, 557, 664   |
| social constructivism .....            | 392, 448, 841   |
| social network analysis.....           | 57, 144, 565, 839, 851, 949, 963  |
| social networking (Facebook, etc)..... | 456, 593, 605, 835, 921   |
| social psychology .....                | 256, 861  |
| surveys and questionnaires .....       | 344, 815, 865   |
| task analysis.....                     | 520   |
| teacher education.....                 | 81, 120, 577, 751, 815, 821, 823, 827, 837, 843, 869, 873, 901, 937, 959, 971   |
| transfer.....                          | 841   |
| university .....                       | 200, 496, 569, 581, 684, 692, 837, 855, 871, 883  |
| video analysis .....                   | 120, 184, 232, 264, 336, 352, 464, 569, 581, 589, 688, 759, 871, 887, 923   |
| videoconferencing .....                | 569, 608  |
| virtual communities .....              | 168, 577  |
| virtual ethnography.....               | 549   |
| virtual/3D environments .....          | 232, 272, 696, 853, 917   |
| visual/performance arts.....           | 280, 700, 899, 953  |
| workplace .....                        | 432, 969  |