

A Mixed-Methods Approach for Studying Collaborative Learning Processes at Individual and Group Levels

Catherine Dornfeld, Naxin Zhao, and Sadhana Puntambekar
cldornfeld@wisc.edu; nzhao23@wisc.edu; puntambekar@education.wisc.edu
University of Wisconsin–Madison

Abstract: Learning processes that unfold during small-group collaboration may impact conceptual outcomes for individual students. To study how learning processes unfolded for eighth grade students collaborating in an e-textbook research activity, we analyzed data sources at individual and group levels using multiple methods, including nonparametric tests, text mining, Markov modeling, and quantitative discourse analysis. Individual measures revealed learning gains on content tests and documentation of shared ideas during collaboration. Group measures revealed increased conceptual discourse over time and streamlining of the research process. Measures at each level indicated distinct paths of inquiry for students and groups; however, these differences were not associated with negative conceptual outcomes. These findings have implications for how we understand how collaborative learning processes unfold, especially between the individual and the group, and how we design support for collaboration and knowledge sharing.

Introduction

CSCL researchers aim to understand both products and processes of learning (Stahl, Koschmann, & Suthers, 2006; Reimann, 2007). Products reveal the relative success of an implementation, but processes reveal the actual learning mechanisms that occur in collaborative settings (Dillenbourg, 1999). To understand collaborative learning processes, we must examine learning at both the individual and group level (Stahl et al., 2006; Dillenbourg, 1999). Individual students contribute unique experiences and prior knowledge, while the group co-constructs knowledge through negotiating and revising shared understandings (Wertsch, 1984).

Understanding learning processes at multiple levels may require mixed-methods approaches, such as quantitative discourse analysis and computational methods (Strijbos & Fischer, 2007; Puntambekar, 2013; Li, Wang, Liao, Zhao, & Huang, 2007). Mixed-methods approaches allow us to combine complementary perspectives and data sources that may triangulate findings about collaborative learning, such as how individuals in groups move toward conceptual convergence (Roschelle, 1992; Kapur, Voiklis, & Kinzer, 2011). We also capture variance in groups' learning processes, which may impact individual students' conceptual outcomes (Barron, 2003). However, with any mixed-methods approach for multiple levels, there are a few major issues. First, in triangulating findings, we must reconcile multiple data sources and identify how each source reveals new understandings of collaborative learning processes (Suthers & Medina, 2011). Also, we must carefully interpret patterns and relationships between data sources (Lajoie, 2011). Finally, we must consider the quality of our data, as poor quality of data can lead to spurious findings (Reimann, Yacef, & Kay, 2011).

In this study, we used a mixed-methods approach incorporating methods from quantitative discourse analysis and learning analytics to understand how learning processes unfolded at both the individual and group levels as students used various forms of support integrated into a curriculum. We analyzed multiple data sources that reflected learning at the beginning and end of an eight-week biology curriculum in order to understand variance between groups and over time. Our research question was: *How do multiple data sources, in combination, reveal how learning processes unfold at both the individual and group levels?* This question has implications for how we understand learning processes using mixed methods for multiple data sources and granularities.

Methods

Here we describe our mixed-methods approach to understanding how different data sources and analyses revealed learning processes at individual and group levels. First, we describe participants and context, then describe the data and connected analyses.

Participants and context

This study focused on three groups of eight-grade students (Groups A, B, and C) working in groups of four ($N = 12$) in the same classroom at a semi-rural public school in the Midwestern U.S. The majority of students

attending this school were Caucasian, and over half of students were eligible for free or reduced-price lunch. Students in the three groups demonstrated similar prior knowledge as assessed on a pre-test.

Students participated in a *Make Your Own Compost* unit. In this eight-week design-based unit, students addressed a challenge about compost and ecosystems. The unit incorporated embedded distributed scaffolds to support students' inquiry, such as physical and virtual experiments; small group collaboration; teacher-led whole class discussions; e-textbook (VidyaMap) research sessions; and scientists' journals for tracking design decisions and relevant content. Physical and virtual experiments supported students' modeling of an authentic problem. Small group collaboration elicited students' current understanding and supported co-construction of knowledge. Whole class discussions facilitated idea sharing between groups and revealed opportunities for teacher support. VidyaMap helped students explore design-relevant content. Lastly, journals included prompts for different aspects of design, such as documenting relevant content.

Several scaffolds intersected during VidyaMap research sessions. These sessions were designed to inform students' design decisions by providing relevant content. In these sessions, students engaged in small-group collaboration to brainstorm research topics; worked in pairs to research topics in VidyaMap using Chromebooks; and recorded individual notes about findings and whole-class ideas in their journals. Here, we emphasize the close interplay between the journal prompts and VidyaMap as scaffolds for learning.

We investigated how each group of students used VidyaMap to research topics at the beginning and end of the unit using journal responses and VidyaMap log data. We also investigated how individual journal responses differed, based on prompts. These prompts guided students to i) brainstorm questions and topics in groups, ii) take notes about research on VidyaMap, and iii) share and record new ideas. We also analyzed group discourse during VidyaMap sessions and content test scores as assessments of conceptual outcomes.

Data and analyses

Here we describe the data sources involved in VidyaMap research sessions and conceptual measures, along with analyses for each data source. Table 1 shows analyses for each unit of analysis. Findings for each analysis are reported in the Results section.

Table 1: Approaches for each unit of analysis

Unit of Analysis	Analytical Approach for Data Source
Individual	<ul style="list-style-type: none"> • Nonparametric tests of content test scores • Topic modeling of journals
Group	<ul style="list-style-type: none"> • Markov models of VidyaMap log data • Small-group discourse analysis

Nonparametric tests of content test scores

To assess learning products at the individual level, we used Wilcoxon Signed-Rank tests to compare students' scores on pre- and post-unit content tests (see Results). Content test focused on concepts related to ecosystems (e.g., biotic and abiotic factors, human impacts, roles and relationships, and cycling of energy and matter). The maximum possible score was 38.5 points, with questions being worth 0.5-3 points depending on question complexity. We also used independent-sample Kruskal-Wallis tests to test for differences between groups at both times. While test scores can demonstrate conceptual outcomes, we considered that the scores do not reveal learning *processes* during the unit. As such, we explored additional analyses to better understand how learning processes unfolded for each group.

Topic modeling of students' journal responses

To assess learning processes at the individual level, we analyzed journal responses associated with the first and last VidyaMap research sessions. The first session (2 days) focused on decomposition factors, while the last session (1 day) focused on ecosystems. Students collaboratively decided on topics for research, but each student recorded questions and notes in their own journal. The journal acted as an artifact of individual learning in that students chose what to record, but it also reflected collaborative discussion of topics – thus showing learning at the *intersection* of the individual and the group.

We used topic modeling to understand how individual students' responses overlapped during group collaboration with VidyaMap. We transcribed and categorized responses within each research session based on the journal prompts. Prompts guided students to i) brainstorm questions and topics, ii) research and take notes

on topics, and iii) share and record new ideas during whole class discussions. We selected responses from the first session (4 responses) and the last session (3 responses) for a total of seven responses per student. We transcribed responses from 159 journals for a total data set of 1113 responses.

We programmed the analysis using Python 2.7 and the *NLTK*, *pickle*, and *gensim* packages. We chose the *term frequency-inverse document frequency* (tf-idf) algorithm because it identifies words that uniquely characterize each individual response relative to the whole data set (Witten, Frank, & Hall, 2011). These characteristic words can be considered relatively rare in that they discriminate that response from others (Witten et al., 2011). We manually identified co-occurrences of these relatively rare words within the three groups to examine how individual responses reflected shared topics of discussion (and potentially knowledge co-construction). This topic modeling procedure revealed overlap in how individual students recorded ideas from their group discussions. However, to understand *how* these ideas were discussed, we need to analyze log data from VidyaMap and collaborative discourse during small-group work.

Markov models of log data

Students' journals indicated how individual students recorded research during group collaboration with VidyaMap. We further investigated students' use of VidyaMap at the group level by analyzing groups' log data from the first and last research sessions. These log data reveal how groups coordinated research activities and how pairs of students within each group navigated through VidyaMap. To clean data, we removed records that involved superficial reading (<10 seconds) unless these records were the first topic in a session or acted as the only connection between prior or subsequent topics. We also removed records that were the only instance for that session. Lastly, we combined data for individual logs with the same name and method of access.

We manually identified concepts that were unique to each group for each session and calculated the number of concepts and time per session. We used Markov models to visualize patterns in how students navigated through VidyaMap. Markov models quantify navigation by showing the probability of moving from one concept to the next, such as from "Compost" to "Ecosystem," within a series of records (Witten et al., 2011). Markov models can also reveal snapshots of groups' activity during each session. For each group, we generated Markov models of their VidyaMap activity for the first and last sessions. We programmed the models using Python 2.7 and *NetworkX* and *matplotlib* packages. While the Markov models revealed how groups researched concepts in VidyaMap, they did not reveal how students discussed these concepts within each session. Thus, we examined students' group discourse during collaborative research.

Discourse analysis of small-group talk

To examine how groups discussed concepts in VidyaMap, we investigated group discourse during the first and last VidyaMap sessions. We coded students' turns of talk for i) conceptual talk, which identified concepts and relationships; ii) procedural talk, which indicated collaborative decisions without explaining concepts or relationships; iii) off-task talk, or (iv) N/A for unclear talk (Cohen's $\kappa = 0.904$; see Dornfeld & Puntambekar, 2016). After coding, we calculated the frequency and proportion of each code within each group's discourse. We used z-score tests of homogeneity to identify significant differences in proportions of talk.

Results

Nonparametric tests of content test scores

Table 2 shows summary statistics for groups' scores. Wilcoxon Signed Rank tests indicated that all students did significantly better on the post-test ($W\text{-value} = 1 < 13, p < 0.05$). To check for group differences, we used Kruskal-Wallis tests to compare mean scores. These tests indicated that groups' scores were not significantly different for the pre-test ($H = 2.202 < 5.692, p > 0.05$) or post-test ($H = 2.375 < 5.692, p > 0.05$). Students appeared to have similar prior knowledge and learning gains. While this indicates the unit supported learning for all students, we added analyses to examine students' *learning processes*.

Table 2: Summary statistics for pre- and post-tests

Group	N	Pre-test			Post-test		
		Mean	SD	Variance	Mean	SD	Variance
Group A	4	29.44	4.49	20.18	32.31	3.40	11.56
Group B	4	30.06	3.24	10.52	32.69	2.64	6.97
Group C	4	26.56	2.17	4.72	30.31	1.21	1.47
Overall	12	28.69	3.49	12.19	31.77	2.58	6.64

Topic modeling of students' journals

Topic modeling revealed relatively rare words within each student's response that discriminated that response against others in the data. Some relatively rare words overlapped for individuals within groups, indicating shared topics of discussion. In Table 3, we list co-occurrences of relatively rare words and their frequency. The most frequent co-occurrences across all groups were *effect/affect* (8 responses), *light* (7 responses), *decomposer* (5 responses), *worms* (4 responses), *temperature* (4 responses), and *helps* (4 responses). Group A showed the least overlap (14 co-occurrences), while Groups B and C showed twice as much overlap (35 and 31 co-occurrences, respectively). Overlap was more frequent during the first session (60 co-occurrences) than the last (20 co-occurrences). Figure 1 also shows that overlap was also more frequent during brainstorming sessions (53 co-occurrences) than research sessions (17 co-occurrences) or whole class discussions (10 co-occurrences).

To summarize, topic modeling revealed how relatively rare words that characterized individual responses (per the tf-idf algorithm) overlapped within each group as shared topics of discussion. We found evidence of overlap within each group; however, Groups B and C demonstrated greater overlap than Group A. While we also found overlap between groups, we found that each group investigated unique topics that other groups did not. Lastly, we see that all groups demonstrated less overlap in the last session compared to the first. To triangulate these patterns at the group level, we next examined the VidyMap log data for each group.

Table 3: Relatively rare word co-occurrences for each group (frequencies in parentheses)

Group	First Session			Last Session		
	Brainstorm	Research	Share	Brainstorm	Research	Share
Group A	Fruit (2) Light (2) Helps (2) Makes (2) Affect/effect (2)	Begin (2)	Affect/ effect (2)	---	---	---
Group B	Moisture (2) Warmer (2) Light (3) Environment (2) Temperature (4) Lower (2) Higher (2) Worms (4)	---	Need (2) Degrees (2)	Flow (3) Together (2)	Plants (2)	Work (2)
Group C	Decomposer (5) Affect/effect (4) Helps (2) Light (2) Factors (2)	Slow (2) Levels (2)	Everything (2)	Life (2)	Depend (2) Components (2) Characteristics (2) Released (2)	---

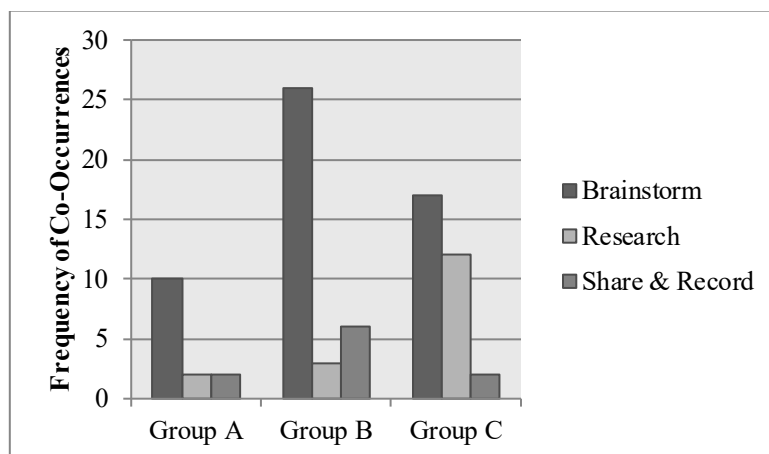


Figure 1. Frequency of word co-occurrences within each group.

Markov models of log data

Analysis of students' journals revealed that each group focused on particular topics of discussion. To triangulate this, we investigated log data from the first and last VidyaMap sessions. The log data revealed that groups read about similar concepts during each session, such as *compost*, *temperature*, and *decomposer*. This makes sense given that groups received the same prompts about decomposition and ecosystems. However, each group's log data also revealed concepts unique to that group. Table 4 lists these concepts.

We found that Group B investigated more unique concepts (12 topics) than Groups A and C (3 and 4 concepts, respectively). On average, Group B investigated more concepts per session (6.7 concepts) than Groups A or C (4.8 and 5.4 concepts, respectively). Group B also spent more time researching (11.9 minutes per session) than Groups A or C (11.2 and 10.7 minutes, respectively). We found that all groups spent less time researching during the last session compared to the first, with a mean session time of 6.2 minutes for the last session and 13.5 minutes for the first.

Table 4: Unique VidyaMap topics in the log data

Session	Group A	Group B	Group C
First Session, Part 1 (Decomposition)	---	Soil Carbon Cycle Water Ecosystems	Nitrogen Consumers Biotic Factors
First Session, Part 2 (Decomposition)	Food Web Biodiversity	Energy Transformation Producers Leaves Stomata Roots Chloroplast	Nitrogen Cycle
Last Session (Ecosystems)	Soil	Food Web Abiotic Factors	---

In Table 5 (next page), we present Markov models that show how students navigated between concepts during each session. These models serve as snapshots of groups' VidyaMap activity that show the probabilities of moving from one concept to the next. Markov models can reveal if students engage in similar or different inquiry, both in terms of content and navigation. We found that the models for the first session were relatively complex compared to the second session. For example, Group C transitions from researching many concepts in the first research session to researching a focused trajectory of concepts during the second session. This decrease in model complexity aligns with the decrease in average session time for all groups. However, these models do not reveal if less model complexity and time spent researching with VidyaMap imply less conceptual discussion of key ideas. Therefore, to understand the focus of students' collaborative discussions with VidyaMap, we used discourse analysis to examine conceptual, procedural, and off-task talk.

Discourse analysis of small-group talk

Discourse analysis revealed that conceptual discourse significantly increased for Groups B and C from the first to last session ($z = 3.35, p < 0.001$; $z = 6.89, p < 0.001$, respectively). In contrast, procedural discourse significantly decreased over time for all groups ($z = 3.09, p = 0.002$; $z = 4.34, p < 0.001$; $z = 8.59, p < 0.001$). Off-task talk did not significantly change over time. Figure 2 (next page) shows these differences. Groups A and B engaged in mostly off-task talk (40.3% and 48.9%, respectively), followed by conceptual talk (35.5% and 32.1%) and procedural talk (24.2% and 19.0%). Group C engaged in mostly conceptual talk (47.3%), followed by procedural talk (35.3%) and off-task talk (17.3%).

Summary of results

Nonparametric tests revealed that students showed conceptual gains on the content post-test. Topic modeling revealed that students working in groups showed overlap in their recorded ideas during VidyaMap research sessions. Markov models of VidyaMap log data also showed overlap in concepts between groups for some concepts, such as *compost* and *decomposer*, but also showed that groups investigated different concepts, such as *biodiversity* and *producers*. Log data also showed that students spent less time researching with VidyaMap during the last session compared to the first, which also aligns with the decrease in model complexity for the last sessions. This decrease in time and model complexity was not concerning, though, as students actually demonstrated more conceptual talk and less procedural talk during the last session compared to the first.

Table 5: Markov models showing topic probabilities for Group B's first and last sessions

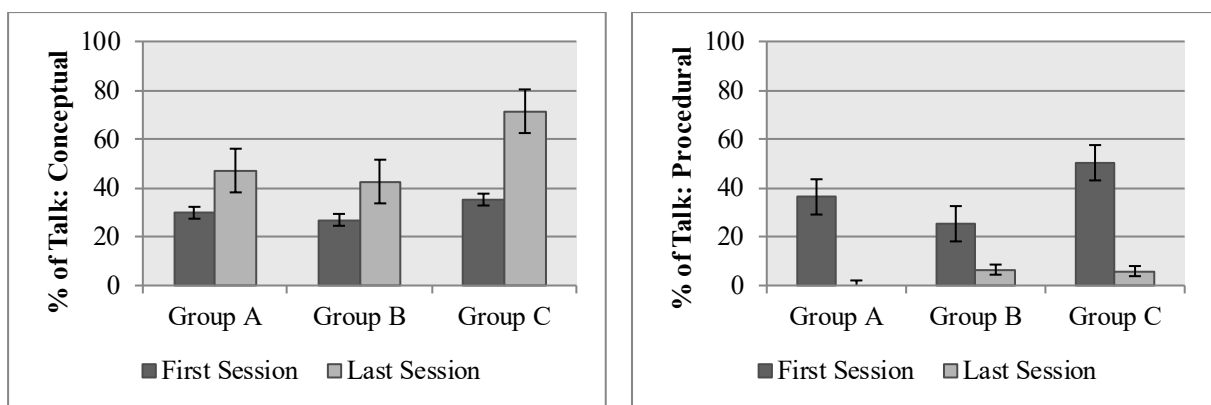
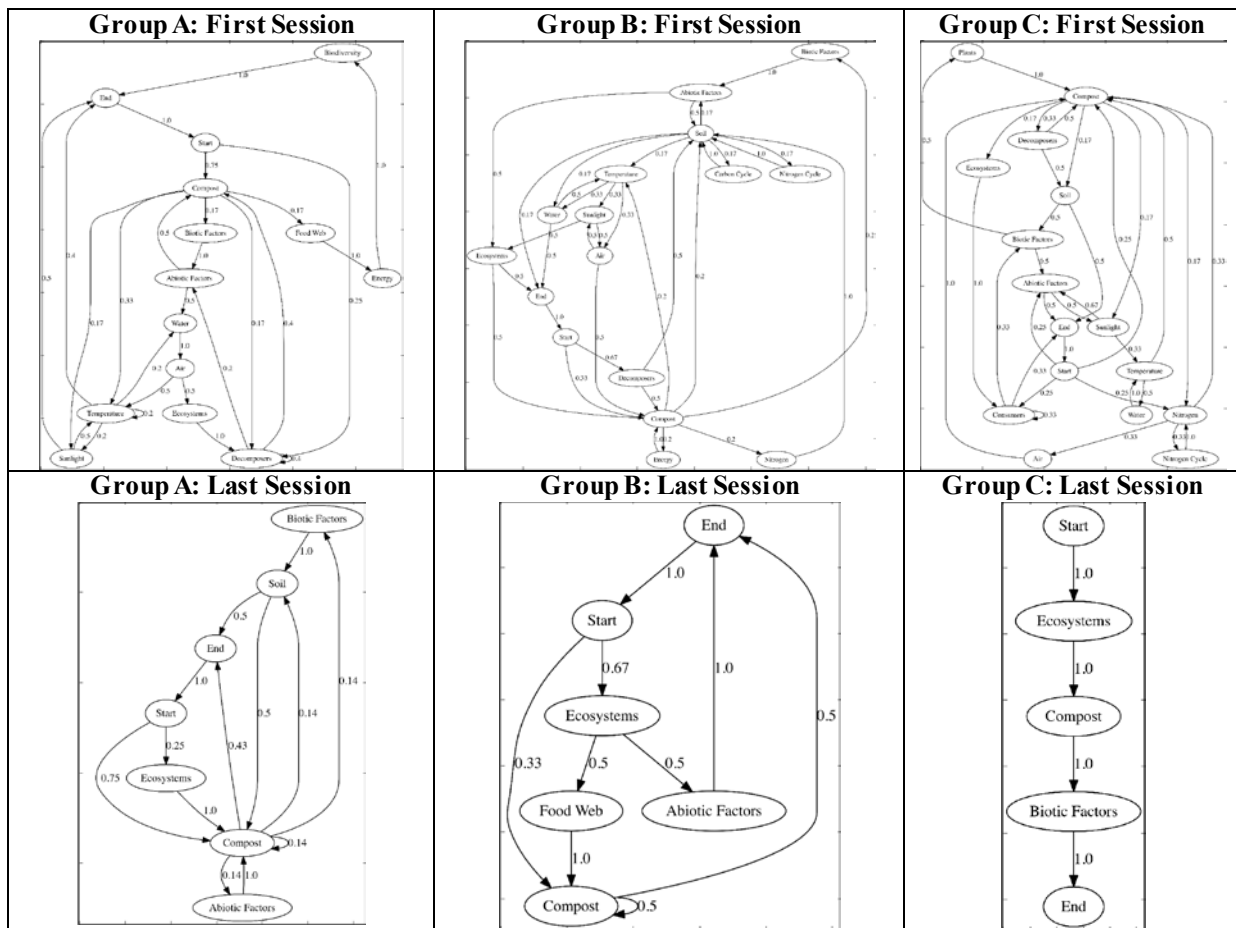


Figure 2. Changes in group discourse over first and last VidyaMap sessions.

Discussion

In this study, we used a mixed-methods approach that incorporated content assessments, topic modeling, Markov models, and quantitative discourse analysis in order to understand the following question: *How do multiple data sources, in combination, reveal how learning processes unfold at both the individual and group levels?* Each analysis addressed a piece of this question, which we summarize and discuss here.

A comparison of learning outcomes showed significantly better performance on the post-test than pre-test along with no differences between groups in pre- or post-test scores, indicating that all students

demonstrated learning gains. However, test scores only give a limited understanding of conceptual outcomes. To understand learning *processes* at the individual and group levels, we investigated student's journal responses to see how individual students working in groups overlapped in their documentation of collaborative research within VidyaMap. Interestingly, when identifying relatively rare words within individual responses, we found overlap among members of the same group, indicating that students discussed and documented shared ideas within their groups. We also found that each group investigated unique topics, based on overlap in journal responses and records from VidyaMap log data. This indicates that each group engaged in collaborative inquiry through distinct paths when using VidyaMap. Even when we detected decreases in average session time and number of topics researched over the unit, we found that this decrease might not be problematic. Students spent less time researching in VidyaMap, but they engaged in more conceptual discourse and less procedural discourse over time. One interpretation of this is that groups used the e-textbook more efficiently to streamline their research, rather than reduce the quality of their research. Groups might have focused more on meanings and applications of concepts instead than procedural decisions about VidyaMap.

Stahl and colleagues (2006), Dillenbourg (1999), and Reimann (2007) have emphasized the importance of studying collaborative learning at both the individual and group levels. By using mixed-methods to study learning processes at both of these levels, we have a better understanding how different groups in the same classroom took different paths to learning, yet arrived at similar conceptual outcomes (Author, 2013; Kapur, Voiklis, & Kinzer, 2011). Using multiple analyses for different data sources allowed us to triangulate how individual students engaged in collaborative discussion of key topics and increased participation in conceptual discourse while keeping track of their own ideas and conclusions (Suthers & Medina, 2011). Examining the relationship between two data sources—the journal and e-textbook—and how they were used in conjunction with each other revealed reciprocal learning processes between the individual and the group; students co-constructed knowledge through brainstorming, researching, and sharing ideas together while individually documenting their ideas. Interestingly, while we found variance in the paths groups took while learning with VidyaMap, students still achieved similar conceptual outcomes. In this study, the variance in content exploration may not have negatively impacted collaboration dynamics (Barron, 2003). Also, the variance may represent a level of tolerance for differences between students in how they individually and collaboratively developed solutions for an open-ended design challenge. Even with different paths toward learning, whole class discussions may have reinforced key ideas between groups. However, for this exploratory study, we cannot be certain of these interpretations without further analysis of whole class discourse (Lajoie, 2011).

To further understand variance between groups and impacts on conceptual outcomes, we plan to investigate embedded opportunities for knowledge sharing during the unit, such as whole class discussions. These discussions facilitate sharing of ideas between groups, which may explain how groups researching different topics demonstrated similar learning gains. These discussions may explain how divergent paths for inquiry are not only tolerable but maybe even helpful if these paths result in greater knowledge sharing, such as with jigsaw activities for knowledge co-construction. As curriculum designers, we may find opportunities for knowledge building between individuals and groups in our design of journal prompts and scaffolding strategies for teachers. We also plan to further investigate how students learn to use VidyaMap as a resource, including streamlining of their research process, by examining their log data and discourse over all sessions in the unit.

The implications of this study involve how we understand collaborative learning processes through combinations of methods that study both the individual and the group. By using multiple methods for each unit of analysis, we better understand how ideas are shared between individuals collaborating in groups, which may impact conceptual outcomes (Barron, 2003). Understanding how conceptual understanding is interwoven between the individual and the group—and also between groups—is essential to our understanding of collaborative learning processes and the design of embedded supports for them.

Conclusion

In this study, we used an exploratory mixed-methods approach to understand how learning processes unfolded at the individual and the group levels during small-group collaboration with an e-textbook. We found that groups engaged in divergent paths of inquiry but still demonstrated similar conceptual outcomes across groups. We plan to further investigate how individuals and groups shared ideas in order to track how groups with divergent paths of inquiry co-constructed shared understandings together. Understanding the progression of knowledge co-construction across the levels of the individual and the group helps us to support collaboration and to track and assess learning outcomes.

References

Barron, B. (2003). When smart groups fail. *Journal of the Learning Sciences*, 12(3), 307-359.

- Dillenbourg, P. (1999). What do you mean by collaborative learning? In P. Dillenbourg (Ed.), *Collaborative-learning: Cognitive and Computational Approaches* (pp. 1-19). Oxford: Elsevier.
- Dornfeld, C. & Puntambekar, S. (2016, June). Negotiation towards intersubjectivity and impacts on conceptual outcomes. In C. Looi, J. Polman, U. Cress, & P. Reimann (Eds.), *Transforming Learning, Empowering Learners: The International Conference of the Learning Sciences (ICLS) 2016, Volume 1* (pp. 562-569). Singapore: The International Society of the Learning Sciences.
- Hutchins, E. (1993). Learning to navigate. In S. Chaiklin & J. Lave (Eds.), *Understanding practice: Perspectives on activity and context* (pp. 35-63). New York: Cambridge University Press.
- Kapur, M., Voiklis, J., & Kinzer, C. K. (2011). A complexity-grounded model for the emergence of convergence in CSCL groups. In S. Puntambekar, G. Erkens, & C. Hmelo-Silver (Eds.), *Analyzing Interactions in CSCL* (pp. 3-23). New York, NY: Springer.
- Lajoie, S. P. (2011). Is the whole greater than the sum of its parts? Explaining the role of individual learning and group processes in CSCL. In S. Puntambekar, G. Erkens, & C. Hmelo-Silver (Eds.), *Analyzing Interactions in CSCL* (pp. 235-246). New York, NY: Springer.
- Li, Y., Wang, J., Liao, J., Zhao, D., & Huang, R. (2007, July). Assessing collaborative process in CSCL with an intelligent content analysis toolkit. In *Seventh IEEE International Conference on Advanced Learning Technologies (ICALT 2007)* (pp. 257-261). IEEE.
- Puntambekar, S. (2013). Mixed methods for analyzing collaborative learning. In C. E. Hmelo-Silver, C. A. Chinn, C. Chan, & A. M. O'Donnell (Eds.), *The International Handbook of Collaborative Learning* (pp. 220-232). New York, NY: Routledge.
- Reimann, P. (2007). Time is precious: Why process analysis is essential for CSCL (and also can help to bridge between experimental and descriptive methods). In C. Chinn, G. Erkens & S. Puntambekar (Eds.), *Minds, Minds, and Society. Proceedings of the Computer-Supported Collaborative Learning Conference (CSCL 2007)* (pp. 598-607). New Brunswick, NJ: International Society of the Learning Sciences.
- Reimann, P., Yacef, K., & Kay, J. (2011). Analyzing collaborative interactions with data mining methods for the benefit of learning. In S. Puntambekar, G. Erkens, & C. Hmelo-Silver (Eds.), *Analyzing Interactions in CSCL* (pp. 161-185). New York, NY: Springer.
- Roschelle, J. (1992). Learning by collaborating: Convergent conceptual change. *Journal of the Learning Sciences*, 2(3), 235-276.
- Stahl, G., Koschmann, T., & Suthers, D. D. (2006). Computer-supported collaborative learning: An historical perspective. In R. K. Sawyer (Ed.), *Cambridge Handbook of the Learning Sciences* (pp. 409-426). Cambridge: Cambridge University Press.
- Strijbos, J. W., & Fischer, F. (2007). Methodological challenges for collaborative learning research. *Learning and Instruction*, 17(4), 389-393.
- Suthers, D. & Medina, R. (2011) Tracing interaction in distributed collaborative learning. In S. Puntambekar, G. Erkens, & C. Hmelo-Silver (Eds.), *Analyzing Interactions in CSCL* (pp. 341-366). New York, NY: Springer.
- Wertsch, J. V. (1984). The zone of proximal development: Some conceptual issues. *New Directions for Child and Adolescent Development*, 1984(23), 7-18.
- Witten, I. H., Frank, E., & Hall, M. A. (2011). *Algorithms: The Basic Methods Data Mining: Practical Machine Learning Tools and Techniques* (3rd ed., pp. 85-146). Burlington, MA: Morgan Kaufmann.

Acknowledgments

We would like to thank Matthew Berland for his assistance with Python. This project received funding from the National Science Foundation (award #1418044).