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) and 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 the 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 ρ(0.65)<0.05; **indicates ρ(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	*"Steel can carry a big load, but it is heavy and weighs down on the recharge interval, and it is a costly option."*	0.89**	0.89*	0.89**
PERFORMANCE PARAMETERS	Attributes: payload, recharge interval, agility, safety, or cost.	*"My device has a pretty good safety, payload, agility, and recharge interval"*	0.89**	1.00**	0.89**
TECHNICAL CONSTRAINTS	Inputs: actuators, ROM, materials, power sources, or sensors.	*"Our two best were both made with Aluminum, NiCd Batteries, Piezoelectric sensors, and Pneumatic actuators."*	0.83**	0.94**	0.89**
CLIENT AND CONSULTANT REQUESTS	Decisions based on internal consultant's requests or client's health or comfort	*"We tried to meet at least the minimum of each of the internal consultant's request."*	1.00**	1.00*	1.00**
COLLAB-ORATION	Facilitating a joint meeting or the production of team design products.	*"How should we make our team batch?"*	1.00**	1.00*	1.00**
DATA	Discussion of numerical values, results tables, graphs, research papers, or relative quantities.	*"I thought that safety near the maximum was not very good (close to 225 - one had 218 RPN)"*	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 $d = 0.57$).

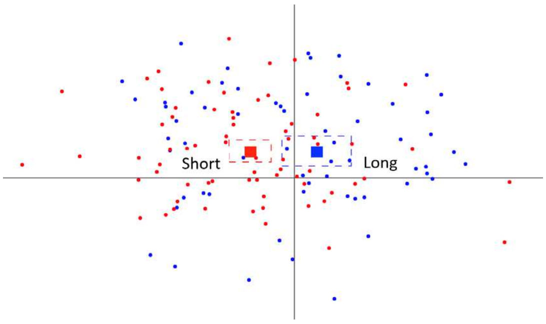

Figure 1. ENA scatter plot showing long (blue) and short (red) readers. Each point is a single student; the squares are group means; the dashed boxes are 95% confidence intervals (t-distribution).

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.

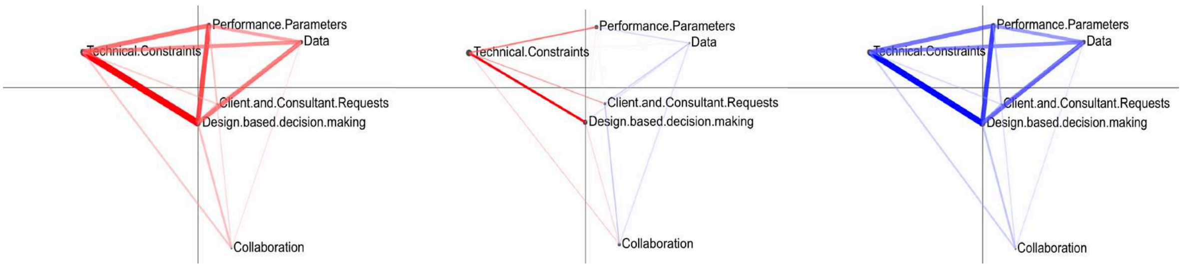

Figure 2. Mean ENA network diagrams showing the connections made by long readers and short readers. Short readers (left, red) mainly connected TECHNICAL CONSTRAINTS with other elements; long readers (right, blue) primarily connected DATA with design reasoning elements.

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 PERFORMANCE PARAMETERS TECHNICAL CONSTRAINTS 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 TECHNICAL CONSTRAINTS 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 PERFORMANCE PARAMETERS TECHNICAL CONSTRAINTS 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 is argument against an alternative to his original proposal.

Comparing ENA models for Mike and Joe

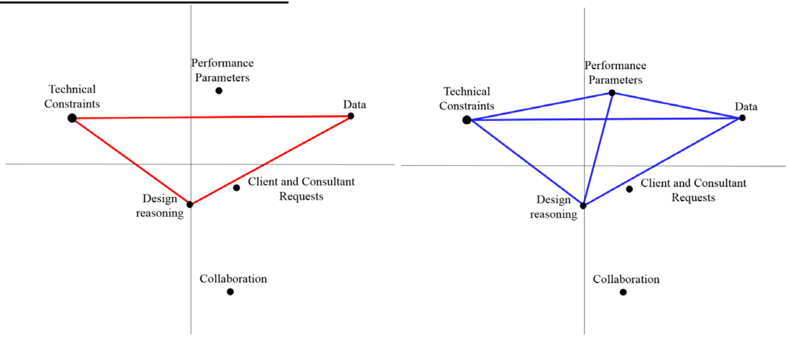

Figure 3. Individual ENA networks showing the connections made by Joe, the long reader (right) and Mike, the short reader (left). Joe connected DATA with TECHNICAL CONSTRAINTS, DESIGN REASONING, and PERFORMANCE PARAMETERS; Mike (left) connected only TECHNICAL CONSTRAINTS with DATA and DESIGN REASONING.

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.