36th annual meeting of the Cognitive Science Society
Quebec City, Canada, July 23-26, 2014

Using Dual Eye-Tracking to Evaluate Students' Collaboration with an Intelligent Tutoring System for Elementary-Level Fractions

Daniel Belenky (dbelenky@andrew.cmu.edu)
Michael Ringenberg (mringenb@cs.cmu.edu)
Jennifer Olsen (jkolsen@cs.cmu.edu)
Vincent Aleven (aleven@cs.cmu.edu)
Carnegie Mellon University - Human-Computer Interaction Institute, Pittsburgh, PA, USA
Nikol Rummel (nikol.rummel@rub.de)
Ruhr-Universität - Institute of Educational Research, Bochum, Germany

Abstract

As learning technologies proliferate, it is important for research to address how to best align instruction to educational goals. For example, recent evidence indicates that working collaboratively may have unique benefits for facilitating the acquisition of conceptual understanding, as opposed to procedural fluency (Mullins, Rummel & Spada, 2011). To investigate this effect, we leverage and expand upon a new methodology, dual eye-tracking, to understand how collaborators' joint attention may impact learning in a collaboration-enabled Intelligent Tutoring System for fractions. We present results from a study in which 28 pairs of 4th and 5th grade students completed a set of either conceptually- or procedurally-oriented instructional activities in a school setting. Results indicate that students collaborating exhibited learning gains for conceptual knowledge, but not for procedural knowledge, and that more joint attention was related to learning gains. These results may inform the design of future learning technologies, and illustrate the utility of using dual eye-tracking to study collaboration.

Keywords: Collaboration; Intelligent Tutoring System, Dual Eye-Tracking; Conceptual Learning.

Introduction

One of the most successful applications of cognitive science to real-world settings has been through the development of Intelligent Tutoring Systems (ITSs). These learning technologies have been shown to help students *learn by doing* as they solve problems by providing targeted feedback in response to errors, as well as next-step hints when students request one. The research presented here is extending these lines of work to allow for pairs of students to collaborate as they engage with an ITS, so students can have the benefits of collaboration while also receiving the cognitive support that ITSs provide. Building on prior research from the field of Computer-Supported Collaborative Learning (CSCL), there is reason to believe that collaborating may be particularly well-suited to facilitate the development of *conceptual* knowledge (Mullins, Rummel & Spada, 2011). Working collaboratively requires students to discuss, mutually elaborate, question, and construct their knowledge, which has been shown to promote a deeper understanding of the materials (see Chi's ICAP hypothesis; Chi, 2009). As robust knowledge consists of both conceptual understanding and procedural fluency, it is important to understand what sorts of instructional environments and activities are better suited towards different learning outcomes.

The current study is leveraging a recent methodological advancement, *dual eye-tracking* (e.g., Jermann, Mullins, Nüssli, & Dillenbourg, 2011), to better understand how collaboration may influence learning. Dual eye-tracking refers to the recording, synching, and analyzing of eye-tracking information from two different students, who, in the present study, worked at two different machines (seeing roughly equivalent interfaces). We use *gaze recurrence* analysis (Richardson and Dale, 2005) to describe, both quantitatively and qualitatively, the different patterns of collaboration engendered by procedural and conceptual learning materials. This analysis method quantifies the degree to which the two collaborators' gazes are in agreement (defined as looking at or near the same place on the interface) at any given point in time, and may provide an index of the quality of interaction (e.g., Nussli, 2011). This data is frequently graphed as a recurrence plot, which provides a way to visualize patterns of joint attention. In the present work, we introduce the methodological contribution of integrating ITS log data into such gaze recurrence plots, and illustrate this method's utility in studying the dynamics of interaction that contribute to successful learning.

Another contribution of the research presented here comes from working with a sample drawn from a much younger population than is generally examined in CSCL research. Working with a younger population in a school-based setting provides an important test of the generalizability of prior findings and theories to a wider range of students and situations. Even with this younger age group, we expect to see that collaboration can help develop conceptual understanding, and that, collaborators can benefit from more conceptually-oriented learning materials, compared to more procedurally-oriented instruction. We test this hypothesis using a collaboration-enabled version of the *Fractions Tutor* (https://mathtutor.web.cmu.edu/info), an ITS that has been shown to produce learning gains for elementary fractions.

The larger goal of our research program is to develop adaptive learning technologies that optimize instruction by matching the type of learning activity with the type of knowledge that is the target of instruction (see the

Knowledge-Learning-Instruction framework, Koedinger, Corbett, & Perfetti, 2012). The research presented here represents a preliminary examination towards this end, focusing on three specific questions: 1) Are 4^{th} and 5^{th} grade students able to show learning gains from a short period of instruction with a collaboration-enabled ITS? 2) Is the development of conceptual knowledge especially facilitated when collaborators work on conceptually-oriented learning materials, compared to procedurally-oriented materials? 3) Is joint visual attention related to increases in learning?

Method

Participants

Participants in this study were 84 4^{th} and 5^{th} grade students from a Western Pennsylvania school district, who participated in 45-minute "pull-out" sessions (in lab rooms set up in their school) during normal instructional time. Their ages ranged from 9-12 years old, M = 9.96, SD = .75. They were assigned to dyads based on their teachers' pairings, and each dyad was randomly assigned to one of four conditions, created by crossing two factors; whether learning was collaborative or individual, and whether the learning materials were geared towards acquiring conceptual knowledge or procedural knowledge. As the present hypotheses are only concerned with the collaborative conditions, the sample of interest here are the 28 students in the collaborative/conceptual and 28 students in the collaborative/procedural conditions (see Olsen, Belenky, Aleven, & Rummel, submitted, for more details on all aspects of the study). The data presented here is at the dyadic level, so that each dyad's joint eye-tracking data can be compared to an average of the dyad's test performances. Learning data from one dyad has been removed, as the post-test data was unusable due to experimenter error, but the eye-tracking data was retained.

Materials

Learning materials. The materials for this study were built using the Cognitive Tutor Authoring Tools (CTAT, freely available from http://ctat.pact.cs.cmu.edu), which have recently been updated to include support for collaborative interaction between two or more students working on the same problem (Olsen et al., submitted). A set of 16 conceptual and a set of 16 procedural learning activities that cover basic fraction equivalence were developed. Each set of 16 consists of four types of problems, with four isomorphs of each type. The materials were sequenced such that students completed one of each of the four types of problems for their condition (procedural or conceptual) before beginning a new set of isomorphic problems. Time-on-task was controlled, with students completing as many of these learning problems as they could in 45 minutes.

The *conceptual* problems focus on understanding underlying principles of fraction equivalence, and how individual components (e.g., numerators, denominators) are interrelated (see Figure 1a). For example, some problems have students compare and contrast two example explanations dealing with whether or not two fractions are equivalent. One of the explanations is correct, but the other reflects a common misconception. Students are tasked with deciding which is correct and why. In another type of problem, students manipulate numerators and denominators of given fractions to see how they relate, and use this information to define what makes fractions equivalent. The *procedural* problems, in contrast, are focused on scaffolding student problem solving as they create and compare equivalent fractions (see Figure 1b). For example, one type of problem has students list the factors of both the numerator and denominator to find the greatest common factor, which is then used to reduce the fraction. Another has students decide if fraction A is equivalent to fraction B by making a series of fractions equivalent to A, and seeing if fraction B is in that list. These problems focus exclusively on the steps needed to complete the procedure, but do not ask the students how or why the procedures work.

The *collaborative* tutors scaffold collaboration by varying problem features available to each partner working on a shared problem. That is, students are given different roles throughout the problems, such as the "problem solver," or the "helper." The problem solver is tasked with inputting responses, based on discussion with her partner. The helper is tasked with aiding her partner in coming up with a correct solution. Students are sometimes given unique information they must share with their partner, creating a sense of individual accountability, where the student must add her voice to the discussion to proceed. All of the various tasks (e.g., solving, sharing, asking) are clearly labeled with appropriate icons (e.g., a "do" icon, a "share" icon, an "ask" icon, etc.). In addition, some steps provide opportunities for group knowledge awareness (Janssen & Bodemer, 2013) by asking each student to first respond independently to a question, and then showing each student's answer to one another. This affords an opportunity for discussion, particularly in cases where there is disagreement, before submitting an answer that is tutored by the system. These features are in addition to other "standard" ITS features that provide cognitive support, such as an interface that breaks problems into steps, targeted feedback, and on-demand hints for each step. Student interactions, like mouse clicks and keyboard entries, are logged by the ITS.

Test materials. A computer-based test was developed to closely match the target knowledge covered in the tutors. The test comprised 5 procedural and 6 conceptual test items, based on pilot studies with similar materials and population. The pre-test was administered in the morning on the day that the student would be using the learning materials, and the post-test was administered the following morning. Students had up to 25 minutes to complete the 11-item test form, and almost all were able to do so. Two isomorphic sets of questions were developed, and there were no differences in performance on these two test forms in the present study, t (79) = .96, p = .338. The presentation of these forms as pre- or post-tests was counter-balanced.

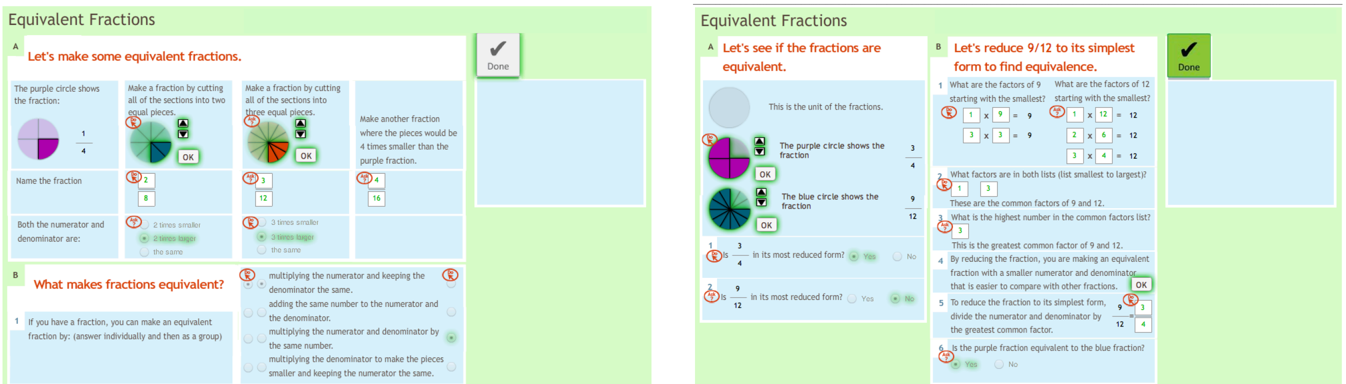


Figure 1a (left) and 1b (right). The first collaborative conceptual (left) learning activity, and the first collaborative procedural (right) learning activity. The collaborative activities stress understanding of underlying principles of fraction equivalence, while procedural activities focus on steps necessary to produce and evaluate equivalent fractions.

Eye-Tracking

Participants completed the learning activities on a 22-inch screen equipped with an SMI Red 250 Hz infrared eye-tracking camera (www.smivision.com). The eye-tracking data generated by the camera includes log messages sent directly from the ITS. As discussed in the introduction, this methodological contribution allows for a tight synchronization between students' observable actions in the tutor interface and their eye-tracking behavior. For example, when students interact with the tutor to input a response, the tutor will immediately evaluate whether it is correct or incorrect, and this can be included in the recurrence plot.

Gaze Recurrence. The gaze recurrence analysis can be conceptualized as asking, "For each two-second slice, what proportion of fixations were at the same location for both students?" This information can be analyzed numerically, as well as displayed graphically in recurrence plots (see Figures 2 and 3). In these plots, if point (t_1, t_2) is dark, it means that at time point t_2, Student 2 fixated on the same screen location on which Student 1 fixated at time point t_1. Our particular focus is on points representing joint attention – that is, when t_1 is equal to t_2 – which are plotted along the diagonal of the recurrence plots. Specifically, gaze recurrence was calculated by first binning the data into two-second slices. As the eye-tracker was sampling at 250 Hz, this provides a maximum of 500 data points for each student for each two-second slice. Considering only fixations (non-fixation data was removed), we calculated, for each two second slice, the proportion of data points in which students' gazes were co-located, defined as being less than 100 pixels apart. The criterion of 100 pixels was chosen because it is similar to what has been used in prior research (i.e., 70 pixels in Jermann et al., 2011), and is close to the size of the interface elements.

Numerical analyses will focus on the *proportion* of data points that indicate joint attention, which we define as when the collaborators are looking in the same area within two seconds of one another. In addition, qualitative analysis of the complete interaction can be examined by graphing the data according to a color scale, with darker colors indicating a larger proportion of fixation-based data points being located in the same area (see Figures 2a, 2b, 3a, and 3b). Dark areas along the diagonal indicate joint attention (i.e., that participants were looking at the same areas of the screen at the same time), while dark points either just above or just below this line indicate that one participant "led" and the other followed his gaze. Dark points further away from the diagonal indicate that a certain area of the screen was fixated by each student but not in close temporal proximity. Location information is not encoded in the plot; dark pixels represent gaze convergence in a certain interface area, but the graph itself does not say which area.

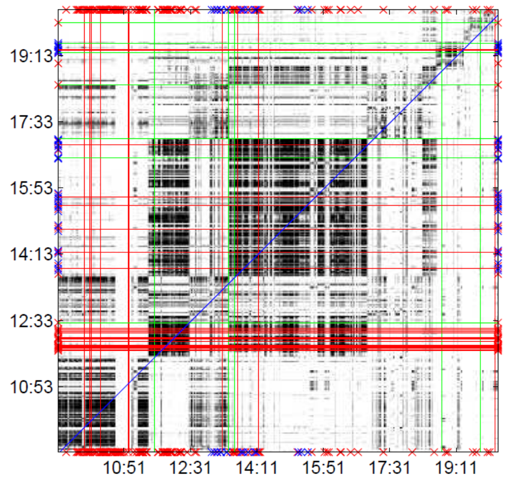
Results

Learning Data

The tutor was effective in helping students gain conceptual knowledge. As revealed in a repeated-measures ANOVA, with pre/post scores on the conceptual test items as the dependent variables, and condition (procedurally- or conceptually-oriented instruction) as a between-subject factor, students increased their conceptual test scores from pre-test (M = 2.06, SD = 1.25) to post-test (M = 2.56, SD = 1.05), F (1, 25) = 7.66, p = .010. However, there was no effect of condition, F (1, 25) = .01, p = .922, nor an interaction, F (1, 25) = .00, p = .99.

There were no differences in a similar analysis comparing procedural test scores on the pre-test (M = .70, SD = .77) to post-test (M = .87, SD = .84), F (1, 25) = 1.13, p = .296. There was, again, no effect of condition, F (1, 25) = .93, p = .345, nor an interaction, F (1, 25) = 1.13, p = .296.

This pattern of results may indicate that, regardless of instructional activity, there is a benefit to collaborating for the development of conceptual understanding, which supports our first hypothesis. However, we do not see evidence that conceptually-oriented instruction facilitates the acquisition of conceptual knowledge more than procedurally-oriented materials do, contrary to the second hypothesis.



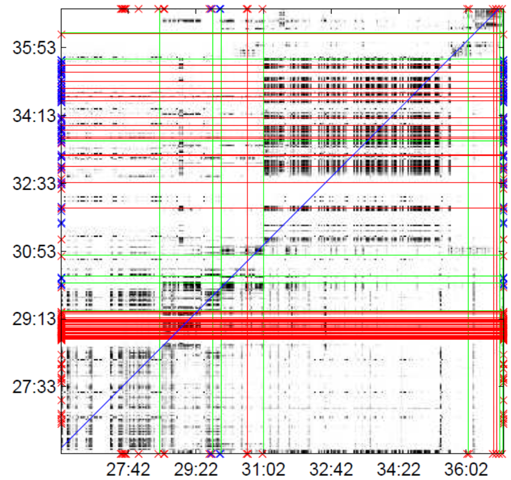


Figure 2a (left) and 2b (right). *Gaze recurrence plots for a high-performing (left) and low-performing (right) conceptual dyad, on the first conceptual problem in the tutor. Darker areas along the diagonal indicate a greater proportion of synchronized gazes. Interaction data from the ITS is overlaid, with red lines indicating moments when incorrect attempts were entered and green lines indicating correct attempts. The axis labels are time stamps for each student.*

Eye-Tracking Data

Joint attention was calculated for each dyad and for each separate problem. Because students completed a variable number of problems, ranging from 2 to 14, (M = 6.96, SD = 2.83), as a first, gross measure, we averaged the joint attention measures for the first four problems (see Table 1). This provides a measure of the amount of time collaborators spent jointly attending to the same information during their first attempt at each of the 4 problem types, which represented the bulk of the 45-minute instruction for most dyads. The reliability of the gaze convergence measure (Cronbach's Alpha, see Table 1) was acceptable across both the conceptual and procedural problems, encouraging given there were only 14 dyads per condition. Thus, there appear to be systematic dyad-level differences; those who had greater gaze convergence on one problem tended to have greater gaze convergence on other problems, inspiring confidence that the gaze recurrence measure captures information about dyads' characteristic patterns of joint attention across problems.

We investigate if this measure of joint attention can be used as an index of the quality of collaboration by analyzing if pairs who more frequently jointly attend to the same information learn more and perform better (Nussli, 2011). To separate out the effect of prior knowledge, gaze was correlated to separate learning gain scores for the procedural and conceptual test subscales, calculated by subtracting pre-test from post-test. The amount of joint attention was not

Table 1. *Means (and standard deviations) proportion of fixations with joint attention for the first four problems.*

	Problem 1	Problem 2	Problem 3	Problem 4	Alpha
Conceptual	.19 (.13)	.13 (.08)	.19 (.11)	.14 (.12)	.75
Procedural	.19 (.11)	.19 (.12)	.21 (.13)	.14 (.10)	.57

correlated to the procedural test gain score, r = .14, p = .491, but there was a marginally significant correlation between joint attention and improvement on the conceptual test, r = .35, p = .072. Interestingly, this effect was localized to the procedural condition, r = .67, p = .012. This correlation was not observed in the conceptual condition, r = .08, p =. 777. Thus, joint attention may have been particularly important for students working on the procedural problems to induce conceptual knowledge, whereas students working on the conceptual problems were able to learn the same information with less joint attention.

Dyadic-Level Comparisons. One approach to understanding *how* collaboration influences outcomes is to compare gaze recurrence plots for high-performing and low-performing dyads. This comparison may provide insight as to how different patterns of interaction are related to different outcomes. It also demonstrates the utility of our novel methodology of overlaying data from the ITS onto the gaze recurrence plot. First, we begin with the conceptual condition, and compare gaze recurrence plots for the first problem for a dyad with a high post-test score to one of the dyads with a low post-test score (see Figures 2a and 2b). We chose the first problem because dyads produced a fair number of errors on this problem, as they were just beginning on the learning activities and were not immediately familiar with how to proceed. In these figures, student behaviors with the tutors are shown, with red lines indicating moments where students inputted an incorrect response, and green lines indicating a correct response.

The two plots show a clear pattern where the high-scoring dyad had, overall, much greater gaze convergence. Specifically, they have more areas with some amount of dark points, indicating more moments with shared attention, and have darker areas, indicating a greater proportion of co-located fixations. The red lines indicate moments when the tutor provided feedback indicating the student response was incorrect, and, as is clear, both groups produced a high

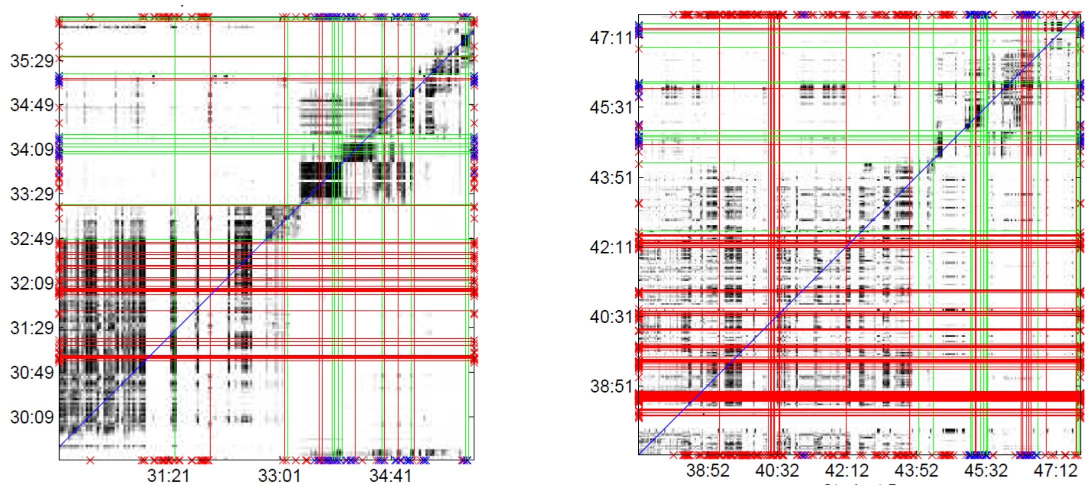


Figure 3a (left) and 3b (right). *Gaze recurrence plots for a high-performing (left) and low-performing (right) procedural dyad, on the first procedural learning problem.*

number of these moments in the early phase of the problem. However, the high-performing dyad maintained a high level of joint attention, as evidenced by a number of incorrect attempts traversing a dark "block" along the diagonal. That is, even as students were entering a number of incorrect responses, they still maintained a high level of joint attention. In contrast, the dyad which ended up doing poorly did not share their gazes very much at all, with only one "block" along the diagonal near the end of the problem.

A similar pattern is observed in Figures 3a and 3b, which show gaze recurrence plots for one high-performing and one low-performing procedural dyad on the first procedural learning problem. Again we see that the high-performing dyad has more areas with a high degree of joint attention (the darker areas). In particular, about halfway through the problem, the high-performing dyad begins a series of periods with intense gaze convergence, and, it is at this point that they begin to input a series of correct responses relatively rapidly (the green lines). The low-performing dyad has a more diffuse pattern of joint attention, and, even when they begin to enter correct responses, their attentional focus does not converge as strongly.

Discussion

We had hypothesized 1) that working collaboratively would produce learning gains, 2) that the conceptual instruction would particularly benefit collaborators, producing greater conceptual learning gains than the procedural instruction, and 3) that increased levels of joint attention would be related to greater learning gains. We will address these hypotheses in order.

Students who collaborated showed learning gains, as predicted. By contrast, students who worked individually on similar problems did not show learning gains on either procedural or conceptual knowledge (we do not focus on this group in the present paper; see Olsen et al., submitted). As such, it appears that building opportunities and support for collaboration can be a beneficial addition to ITSs.

We had expected that the conceptually-oriented instruction would produce higher conceptual learning gains for collaborators, compared to the procedurally-oriented instruction. Evidence for this prediction was not observed. The absence of this effect may be due to a small number of methodological factors. First, it may be that the short duration of the instruction (45 minutes) lowered the likelihood for beneficial interactions to emerge. In particular, students working collaboratively completed an average of 6.96 problems, compared to an average of 10.41 among students working by themselves. However, the findings that students in the collaborative conditions showed learning gains, while those who completed more problems individually did not, indicates the potential effectiveness of having students collaborate. Finally, the test items may not be sensitive to all forms of learning that may have occurred. While the test items were closely aligned to the instruction, other measures of transfer, such as preparation for future learning (Schwartz & Martin, 2004), may have revealed longer-term benefits for the conceptual instruction. Given these constraints, it is particularly encouraging that we found evidence that even elementary-school students can productively collaborate while using an ITS, and that this was observed in an ecologically-valid school setting.

Turning to the dual eye-tracking data, we observed that there were reliable between-dyad differences in joint attention. We also observed that joint attention was related to learning gains in conceptual knowledge, although only in the procedural condition, a surprising finding. It is possible that, for this condition, only those dyads who actively and constructively engaged were able to induce the underlying conceptual knowledge. This finding suggests that one route to successful conceptual learning may be to have collaborators work on explaining procedures to one another, an intriguing possibility that warrants further investigation.

However, this result also requires considering why joint attention was not related to learning gains for the conceptual condition. One possibility is that joint visual attention was less important for learning from the conceptually-oriented problems, as the more abstract nature of the instruction required engagement with the underlying principles, regardless of where the students were looking at any given moment. This interpretation is supported by the lack of differences in learning between the conceptual and procedural conditions, which indicates that the conceptual condition learned just as well, regardless of joint visual attention. However, it is also possible that this effect stems from differences in the collaborative features of the particular problems. Some of the conceptual problems required students to verbally convey unique information that their partner could not see, which may have reduced the possibility for joint visual attention to emerge. Future research could investigate how particular collaborative features influence joint attention, as well as comparing visual attention with other measures of synchronized attention (e.g., frequent turn-taking in dialogue), to see how each of these are related to successful learning outcomes for different instructional activities.

More broadly, we have attempted to illustrate how dual eye-tracking might be useful for a number of applications, such as the iterative design of successful learning technologies. For example, we observed variability in the amount of joint attention maintained during periods of difficulty on early steps in the conceptual problems, indicating a potential target for additional scaffolding. One possibility would be to develop targeted feedback or highlighting on the tutor interface to guide both students to attend to the same information in response to errors. For example, the helper may be given a prompt that explicitly provides some concrete steps they can take to help the problem solver. Another possibility is to integrate information about the collaborator's current visual position, helping students maintain joint attention (see Schneider & Pea, 2013). Another use of dual eye-tracking is to test hypotheses about patterns of interaction. This was not explored in the present paper, although we did observe that joint attention was consistent for dyads across problems, indicating its potential utility as a marker of collaboration quality. We believe that the methodological contribution of integrating data from the ITS directly into the eye-tracking log will greatly contribute to this sort of research, as this information can be combined with other streams of data (like transcripts and videos of the interaction), helping researchers study the dynamics of productive collaboration.

In the present paper, we have introduced a collaboration-enabled ITS for teaching fractions, and illustrated its efficacy with a short, school-based experiment. Specifically, we demonstrated that having students collaborate leads to increases in conceptual understanding of the materials. In addition, we used dual eye-tracking measures to understand how joint visual attention was related to learning from conceptually-oriented and procedurally-oriented materials, introducing the novel contribution of integrating information from the ITS log with a gaze recurrence plot. Dual eye-tracking is emerging as a useful contributor to the measurement, study, and creation of novel and effective CSCL systems (e.g., Schneider & Pea, 2013). By integrating theories of learning from cognitive science with insights into the dynamics of collaboration revealed by these new data streams, our understanding of collaborative learning, and the technologies to support it, may continue to improve.

Acknowledgments

The work is supported by award R305A120734 from the US Department of Education, Institute of Education Sciences.

References

Chi, M. T. H. (2009). Active-Constructive-Interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science, 1,* 73-105.

Janssen, J., & Bodemer, D. (2013). Coordinated computer-supported collaborative learning: Awareness and awareness tools. *Educational Psychologist, 48* (1), 40-55.

Jermann, P., Mullins, D., Nüssli, M.-A., & Dillenbourg, P. (2011). Collaborative Gaze Footprints: Correlates of Interaction Quality. In CSCL '11: *Proceedings of the 2011 conference on Computer support for collaborative learning.* International Society of the Learning Sciences, 2011.

Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012) The Knowledge-Learning-Instruction framework: Bridging the science-practice chasm to enhance robust student learning. *Cognitive Science, 36,* 757-798.

Mullins, D., Rummel, N., & Spada, H. (2011). Are two heads always better than one? Differential effects of collaboration on students' computer-supported learning in mathematics. *International Journal of Computer-Supported Collaborative Learning, 6,* 421-443.

Nüssli, M.-A. (2011). *Dual eye-tracking methods for the study of remote collaborative problem solving.* (Doctoral dissertation). doi: 10.5075/epfl-thesis-5232.

Olsen, J., Belenky, D. M., Aleven, V., & Rummel, N. (submitted). Using an intelligent tutoring system to support collaborative as well as individual learning. Paper submitted to the *12th International Conference on Intelligent Tutoring Systems*. Honolulu, Hawaii.

Olsen, J., Belenky, D. M., Aleven, V., Rummel, N., Sewall, J., & Ringenberg, M. (submitted). Authoring tools for collaborative intelligent tutoring system environments. Paper submitted to the *12th International Conference on Intelligent Tutoring Systems*. Honolulu, Hawaii.

Richardson, D. C., & Dale, R. (2005). Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science, 29,* 1045-1060.

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.

Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognitive and Instruction, 22* (2), 129-184.