Combining Gaze, Dialogue, and Action from a Collaborative Intelligent Tutoring System to Inform Student Learning Processes

Jennifer K. Olsen, Human-Computer Interaction Institute, Carnegie Mellon University; Ecole Polytechnique Fédérale de Lausanne, jennifer.olsen@epfl.ch
Kshitij Sharma, Norwegian University of Science and Technology, kshitij.sharma@ntnu.no
Vincent Aleven, Human-Computer Interaction Institute, Carnegie Mellon University, aleven@cs.cmu.edu
Nikol Rummel, Human-Computer Interaction Institute, Carnegie Mellon University; Institute of Educational Research, Ruhr-Universität Bochum, nikol.rummel@rub.de

Abstract: In a computer supported collaborative learning environment, students have both interactions with each other as well as the technology that is guiding their learning, which can influence how the students construct their knowledge. Often in technology enhanced learning situations, information from the system provides discrete data points that can be used to infer learning without providing much information on the knowledge construction. On the other hand, analysis of student dialogues can be time consuming and subjective. In this paper, we propose combining log data, student dialogue, and gaze analysis to provide a clearer picture of how students construct knowledge collaboratively while working with an intelligent tutoring system. We found that students' gaze similarity is negatively correlated with levels of abstraction in speech and that students have higher gaze similarity surrounding feedback provided by the tutor. These results show that the gaze data can be used as a proxy for dialogue in a collaborative learning context.

Keywords: Dual eye-tracking, DUET, Intelligent tutoring systems, ITS, gaze analysis, eye-tracking

Introduction

To develop a better understanding of how students construct their knowledge, it is important to be able to efficiently analyze the process that the students go through as they solve them problem. When students are working collaboratively, this process involves both their interactions with the other students in the group as well as the students' interactions with the learning materials. To fully understand how students construct their knowledge, it is important to understand both of these interactions, which cannot always be captured through a single type of process data. When using technology enhanced learning, the data about the student processes is limited to discrete moments of time that can be captured within the logs of the technology. On the other hand, within computer supported collaborative learning, we often have a continuous stream of process data from student dialogues. By combining these multiple data streams, we can develop a better understanding of the student's knowledge construction and how the students' interactions unfold between one another and in relation to the educational material with which they are interacting. Intelligent tutoring systems (ITSs) provide an ideal environment for investigating the relation between these multiple data streams. However, dialogue data can be complicated to analyze and the findings are often very subjective and depend upon the coding scheme that is used.

On the other hand, research shows eye gaze is tied to communication, making eye-tracking a promising method to use as a proxy for dialogue (Meyer, Sleiderink, & Levelt, 1998). Previous research has shown a link between speech and eye gaze that goes in both directions: eye gaze can precede the mention of an object or follow it (Griffin & Bock, 2000; Meyer, Sleiderink, & Levelt, 1998). This same pattern occurs when people work on a task together. There is a coupling of the collaborators' eye gaze around a reference (Richardson, Dale, & Kirkham, 2007), meaning that the collaborators' gaze may fixate, at approximately the same point in time, at the object referenced in the dialogue, for example just before mentioning it and just after hearing about it. The eye gaze has a closer coupling when each of the collaborators has the same initial information and when collaborators can visually share important objects that they are referencing in speech (Jermann & Nüssli, 2012; Richardson, Dale, & Kirkham, 2007), suggesting that concrete references may have more of an impact on eye gaze compared to abstract references.

Over the past few years, eye-tracking has become a key source of process data in educational research. Research using eye-tracking covers a wide range of educational ecosystems. Eye-tracking has multi-faceted use case examples: From online (Sharma et. al., 2015a) to face to face classes (Raca & Dillenbourg, 2013), from co-located (Schneider et. al, 2017) to remote collaborative learning (Sharma et. al., 2015b), and to understand teachers' classroom orchestration processes (Prieto et. al., 2015). Eye-tracking has not only been used to understand the learning processes in various contexts, but it also has been used to provide students appropriate, real-time, and adaptive feedback on their learning processes (Sharma et.al, 2016, D'Angelo et. al., 2017).

In terms of collaborative learning scenarios, eye-tracking has most often been used with collaborating partners' dialogues. Griffin & Bock (2000) showed that there is a time lag of about 800 milliseconds between looking at an object and referring to the same object (eye-voice span). Allopenna et. al. (1998) showed that there is a time lag (about 400 milliseconds) between a speaker's reference and a listener's gaze on the referred object (voice-eye span). In a dual-eye-tracking study, Richardson & Dale (2005) gave the notion of eye-eye (speaker's eye listener's eye) span as the time difference between the moment a speaker looks at an object and the moment the listener looks at the same object. Richardson & Dale (2005) found this lag to be about 1.2 seconds. Most of the dual eye-tracking studies have shown that the amount of time that the collaborating partners spend while looking at the same objects at the same time (cross-recurrence) is predictive of several collaborative constructs (e.g., collaboration quality Jermann & Nuessli, 2012; misunderstandings Cherubini et. al., 2007; learning gains Sangin et. al., 2007). These studies consider the dialogue as basic utterances (Sangin et. al., 2007), referencing words (Jermann & Nuessli, 2012), or in a few cases, as a collaboration quality category (Schneider et. al., 2013). In this paper, we present a new dialogue coding scheme, which captures the abstraction in the dialogue; that is, how much context dependency (low abstraction) or domain knowledge (high abstraction) is reflected in the speech. Moreover, we present the relation between the similarity of the gaze patterns and the level of abstraction in the problem-solving processes.

ITSs have been very successful in supporting students' learning as they work individually to solve problems (Murray, 2003), particularly within the domain of mathematics (Ritter, Anderson, Koedinger, & Corbett, 2007). ITSs are beneficial to students by providing them with cognitive support as they solve a problem (VanLehn, 2011). ITSs provide step-by-step guidance for students both through the use of immediate feedback on steps and through on demand hints. That is, students will know right away when an error occurs and they can decide to request help from the system to figure out how to do any problem-solving step correctly. Although the majority of ITSs have been developed for individual learning, there has been some work combining ITSs and collaboration successfully (Baghaei, Mitrovic, & Irwin, 2007; Walker et al., 2009; Diziol, Walker, Rummel, & Koedinger, 2010). By combining collaboration, which supports learning through processes such as co-construction and explanation-giving (Chi & Wylie, 2014), with the cognitive support provided in the ITS, students may be able to more effectively construct knowledge to both avoid errors, over-come errors when they occur, and to effectively use the support provided through hints. However, it is still an open question about how the events that occur within the ITS impacts the collaboration. By combining gaze data with tutor log data and student dialogues, we can construct a more complete picture as to how students are constructing knowledge while working on the tutor.

Eye-tracking in ITSs has been previously used to better understand student processes during the learning process but has primarily been used to investigate students working independently. The use of eye-tracking as an analysis tool in ITSs has spanned investigating both affective and cognitive states of students (Jaques, Conati, Harley, and Azevedo, 2014; Rau, Michaelis, & Fay, 2015). Within the affective states, eye tracking can be used to gauge student states around boredom, curiosity, and attention that can influence the student learning (Jaques et al., 2014). By identifying these states, interventions can be put in place. In addition to tracking the affective state of the student, the eye gaze can also be related to the cognitive state of the student. Rau et al. (2015) found that the gaze patterns of students were correlated with the types of self-explanations that students provided. However, the majority of this research does not extend the analysis of eye-tracking to students working collaboratively (Belenky, Ringenberg, Olsen, Aleven, & Rummel, 2014). When students work collaboratively, they can influence each other's thought processes that can be expressed through both speech and gaze patterns.

In this paper, we aim to answer two main research questions: what is the relation between collaborative gaze patterns and the level of abstraction in student speech and what is the relation between tutor events and gaze patterns? To answers these questions, we analyzed multiple data streams from elementary school students working with a collaborative fractions ITS including gaze data, tutor log data, and transcript data. We hypothesized that dyads with a higher similarity of gaze data would have a lower level of abstraction in their speech (H1), as students that are talking about specific features within the problem would more likely be looking at the same thing. Second, we hypothesized that gaze similarity would be greater around correct actions, as students would be working together to solve the problem (H2). In the following sections, we will present our three data sources as well as the results from triangulating this data. This paper provides a deeper understanding of how system information can influence and interact with students' collaborations.

Methods

Experimental design and procedure

Our data set involves 14 4th and 14 5th grade dyads from a larger study that tested the hypothesis about differential benefits of collaborative versus individual learning (Olsen, Belenky, Aleven, & Rummel, 2014). Each teacher paired the students participating in the study based on students who would work well together and had similar, but not equivalent, math abilities. The dyads were engaged in a problem-solving activity using a networked collaborative ITS, which allowed them to synchronously work in a shared problem space where they could see each other's actions while sitting at their own computers. The students were able to communicate verbally through a Skype connection. Each dyad worked with the tutor for 45 minutes in a pull-out study design at their school. The morning before working with the tutor and the morning after working with the tutor, students were given 25 minutes to complete a pretest or posttest individually on the computer to assess their learning. During the experiment, dual eye tracking data, dialogue data, and tutor log data in addition to the pretest and posttest measures were collected. We collected eye-tracking data using two SMI Red 250 Hz infrared eye-tracking cameras.

Intelligent tutoring system

During the study, the dyads engaged with an ITS oriented towards supporting the acquisition of knowledge about fraction equivalence. Within each problem, the tutor provided standard ITS support, such as prompts for steps (i.e., revealing steps sequentially), next-step hints, and step-level feedback (i.e., correct or incorrect feedback) that allows the problem to adapt to the student's problem-solving strategy (VanLehn, 2011). Each of these different supports were displayed as actions on the screen that could guide the students' actions and gaze.

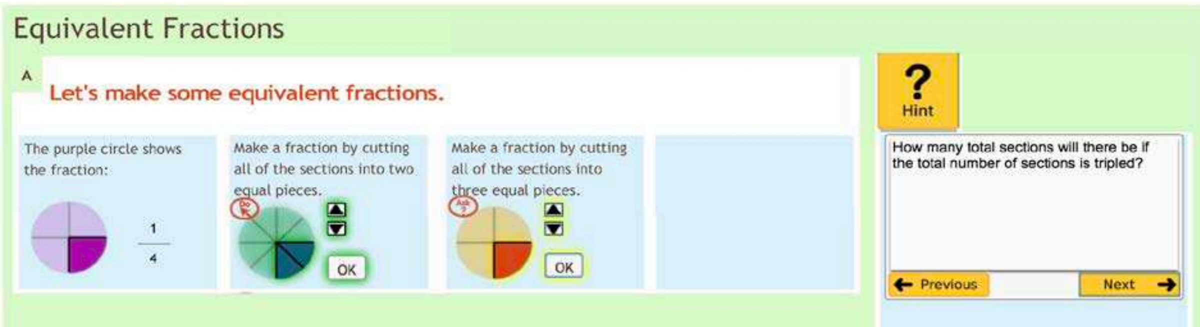

Figure 1. Example of a fractions interface showing incremental step reveals, step feedback, and hint requests. Students had roles assigned by step that were displayed through their icon.

For the collaboration, the ITS support mentioned above was combined with embedded collaboration scripts, which allowed students to take slightly different actions and see different information (see Figure 1). The embedded collaboration scripts included three theoretically proven types of collaboration support: roles, cognitive group awareness, and individual accountability. First, for many steps, the students were assigned *roles* (King, 1999). In the tutors, on steps with roles, one student was responsible for entering the answer and the other was responsible for asking questions of their partner and providing help with the answer. The tutor indicated the current role for the students through the use of icons on the screen. A second way in collaboration was supported was by providing students with information their partner did not have that they were responsible for sharing for the problem to be completed causing *individual accountability* (Slavin, 1996). The final feature was *cognitive group awareness,* where knowledge that each student has in the group is made known to the group (Janssen & Bodemer, 2013). On steps where this feature was implemented, each student was given an opportunity to answer a question individually before the students were shown each other's answers and asked to provide a consensus answer.

Dependent measures

For our analysis, we collected data from three different data streams that were used for analysis: log data, student dialogue, and gaze data. Although the log data recorded each transaction that the student took within the tutor, we were interested in the transactions that ended in additional changes to the tutor interface besides the students own actions. These fell into three categories of *hint requests*, *incorrect feedback*, and *correct feedback.* On each step, the students could request hints from the tutor related to the current step that they were working on. When submitting an answer to the tutor, the students would get either correct or incorrect feedback for each step and would need to have the step be marked as correct before being able to continue with the problem. The log data captured each of these transactions along with a time-stamp.

Each of the student dialogues were transcribed and coded for abstraction levels. Abstraction is how grounded within the concrete aspects of the problem solving and communication the student's utterance is. The level of abstraction is fully dependent on what occurs in the dialogue and is not intended to infer all mental processes. Within our transcripts, we coded for abstraction at the utterance level. This allowed us to have a finer-grained coding for each second of the dialogue without losing the context of the words. The abstraction codes consisted of six different levels: acknowledgement, read out loud, interface, problem solving, concepts, and metacognitive (See Table 1). The levels of abstraction followed an ordering with acknowledgments being the least abstract and metacognitive being the most abstract (with the other codes following the ordering in Table 1). For the coding, all statements that were off-task or were with a researcher were marked as "not applicable" and were discarded from the analysis. An inter-rater reliability analysis was performed to determine consistency among raters (Kappa= 0.78).

Table 1: Abstraction feature coding of student utterances in increasing order of level of abstraction

Code	Description
Not Applicable	The student engages in off-task behavior, converses with the experimenter, or vocalizations without any context.
Acknowledgments	The student acknowledges their partner, or they request acknowledgment or a repeat of what the partner has said.
Read Out Loud	The student is reading information provided within the problem and presented on the screen.
Interface	The student discusses actions that can be taken in the interface or engage in work coordination.
Problem Solving	The student is providing an answer to the problem or showing evidence of think aloud as they solve the problem.
Concepts	The student is adding information from outside of the problem or providing an explanation that goes beyond the required answer.
Metacognitive	The student verbally expressing their understanding of their current knowledge/problem solving state.

To compute the entropy, we divided the screen in 50-by-50 pixels grid. We also divided the whole problem-solving session into 10 seconds time windows. We then computed the proportion of the time spent in each block in the spatial grid for each 10-second time window. This resulted in a series of 2-dimensional proportionality vectors. Finally, we computed the Shannon Entropy for each of the vectors. A low entropy value (the minimum possible value is zero) depicts that the student was looking at only a few elements on the screen, which we called *focused gaze*. On the other hand, a high value of entropy indicates more elements being looked at in a given time window, which we called *unfocused gaze*. Although focus and attention are related concepts, focus, as we defined here, does not contain the idea of processing the stimulus, as is required in the definition of attention. Focused gaze simply indicates a small number of elements looked over a fixed time period.

In order to compute the similarity between the gaze patterns of the collaborating students, we divided the screen space and the interaction time in the same manner as we did for entropy computation. We computed the similarity between the two proportionality vectors by using the reverse function (1/(1+x)) of the correlation matrix of the two vectors. A similarity value of one will show no similarity between the two gaze patterns during a given time window. On the other hand, a higher value of similarity will show that the two participants spent time looking at the similar set of object on the screen during the same time window.

Results

Gaze and tutor response

We compared gaze similarity across time (± 5 seconds) for different types of tutor responses. For this comparison, we fit a hierarchical linear model with time and tutor response as random effects and gaze similarity as the dependent measure. We observe a significant effect showing that similarity values are different among the three types of tutor responses, $F(2, 16.96) = 47.80$, $p < .001$, while there was no significant main effect of time, $F(1, 15.41) = 2.60$, $p = .12$ or interaction between time and tutor response, $F(2, 18.09) = 0.38$, $p = .68$ (see figure 2). A post-hoc pairwise comparison shows that the similarity is the highest for the correct feedback and the lowest for the hint requests (see Table 3).

Table 3: Pairwise comparison of tutor response with gaze similarity

	Correct Feedback	Incorrect Feedback
Hint Request	$F(1,16.41) = 85.96$, $p < .001$ Correct feedback > Hint request	$F(1,16.77) = 23.75$, $p < .001$ Incorrect feedback > Hint request
Correct Feedback	-	$F(1,15.17) = 48.88$, $p < .001$ Incorrect feedback < Correct feedback

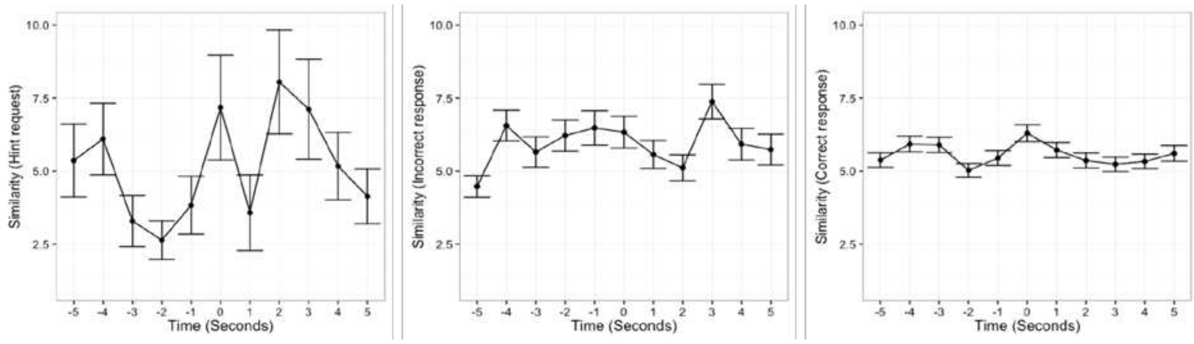

Figure 2. Gaze similarity across time for hints (left), correct responses (center), and incorrect responses (right).

Further, we compared the gaze similarity values across different types of tutor responses and the probability of both students being focused. Again, we fit a hierarchical linear model. We observed a significant effect that as the probability of the students being focused increases, the similarity values increase, $F(1, 19) = 29.37$, $p < .001$ (see Figure 3). Additionally, we observed a significant interaction between the student focus and tutor response type, $F(3, 18) = 6.71$, $p < .05$, with the correlation being greatest for incorrect responses and lowest for correct responses and a marginal significant effect of tutor response on similarity, $F(2, 17) = 2.86$, $p = .06$.

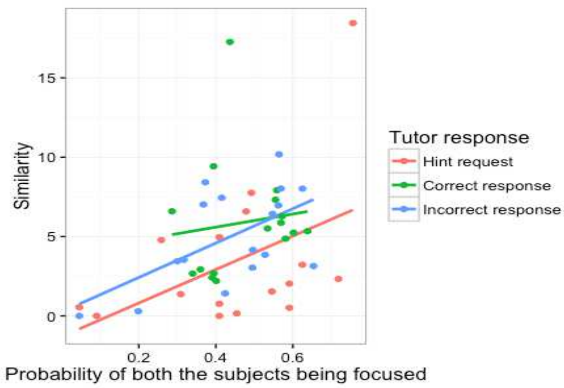

Figure 3. Correlation of gaze similarity with probability of dual dyad focus by tutor response type.

Gaze and abstraction

For the abstraction categories, we observed that the concept category was coded for less than half of the dyads (ten out of 28) and that for these dyads, the concept category was less than 5% of the utterances. Given this low rate, we determined that the concept category would not be a reliable category. Therefore, we merged the concept and problem solving cateories for this analysis.

Table 3: Model parameters for correlation of focus and tutor response with gaze similarity

	Read out Loud	Interface	Problem Solving	Metacognitive
Acknowledgements	N.S.	N.S.	N.S.	$F(1,40.62)=3.19$, $p=.08$ Ack. > Metacognitive
Read out Loud	-	N.S.	$F(1,30.23)=7.98$, $p<.05$ Read out loud > Prob. Sol.	$F(1,41.76)=13.91$, $p<.001$ Read out loud > Meta.
Interface	-	-	N.S.	$F(1,40.96)=6.94$, $p<.05$ Interface > Metacognitive
Problem Solving	-	-	-	$F(1,30.09)=3.69$, $p=.06$ Prob. Sol. > Meta.

To investigate the relation between gaze similarity and dialogue abstraction, we ran a one-way independent ANOVA without assuming equal variances across the different abstraction categories. There was a significant effect of dialogue abstraction with gaze similarity, $F(4, 51.95) = 3.86$, $p < .05$. In a post-hoc pairwise comparison (see Table 3), we found that gaze similarity is significantly lower when students utterances are coded as metacognitive. Additionally, we observed a trend that higher abstraction levels tend to have lower similarity values with an exception of acknowledgments (see Figure 4).

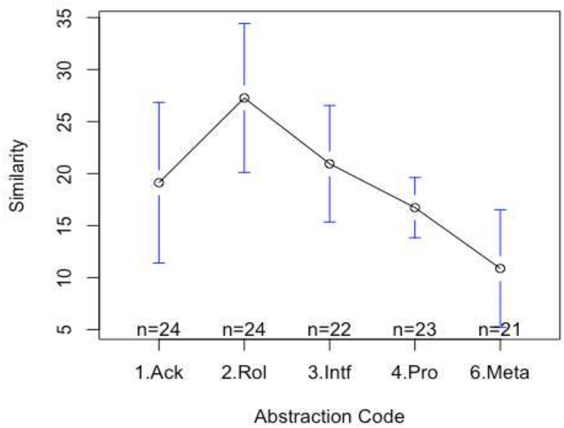

Figure 4. Gaze similarity by dialogue abstraction codes.

Tutor response and abstraction

Finally, we ran a chi-square test to compare the tutor response types with the dialogue abstraction points (for a range of ± 3 seconds around the tutor response). For the different levels of abstraction and tutor responses, we found a significant relation, $\chi^2(8) = 15.61$, $p = .04$. However, Comparing the residuals from the chi-square fit, we did not find any significant differences.

Discussion and conclusions

In this paper, we presented a dual eye-tracking study within a remote collaborative setting where pairs of students solved fractions problems working with an ITS. We combined the gaze and the dialogues for each dyad with their interactions within the ITS. We observed three main relations: 1) the gaze similarity decreases as the level of abstraction in the dialogue increases; 2) the probability of the pair being focused is correlated to the pair's gaze similarity when they receive a hint or an incorrect response; 3) the gaze similarity is highest for the correct response, followed by the similarity for incorrect response and it is lowest for the hint requests.

In respect to the relation between the tutor response and similarity, we observed that for the correct feedback, the similarity is highest. This might be because when students working together, they will be looking at the same thing (leading to a higher similarity) and have a higher chance of getting the solution correct. On the other hand, the similarity is lowest for the hint requests; one of the reasons for this could be that the hint message is displayed above the hint button and there could be a lag of gaze on the hint message between the student who requests the hint and their partner. The similarity for the incorrect feedback is lower than that for the correct feedback. One possible explanation could come from the fact that students not working together are not sharing their knowledge so have less of a chance of getting the solution correct. Another reason could be the fact that once the students receive incorrect feedback they scan the interface for the mistake/correct answer and hence have a lower similarity.

Further, we also analyzed the relation between the tutor response, similarity, and focus of the two students. We found that the focus is correlated with the similarity, however this correlation is highest for the incorrect feedback and lowest for the correct feedback. Related to what we found with the similarity and tutor actions, this observation could be due to the fact that once the students receive incorrect feedback, they start by scanning the interface for the mistake/correct result and then they focus together on the same elements of the problem, which increases their similarity values. On the other hand, once the students receive correct feedback, they start by working independently on the next problem. This makes the probability of them being focused at the same time and looking at the same part of the screen almost independent from each other.

Additionally, we observed that the similarity decreases as the level of abstraction in the dialogue increases. A plausible explanation could be that the support required from the stimulus decreases with an increase in the level of abstraction because as students use higher level of abstraction they use a fewer number of physical (on visual stimulus) references. As an example, during a read-out-loud dialogue students are reading from the screen (a visual stimulus) and hence they have a high similarity; while during a metacognitive dialogue, they do not refer to anything present on the screen, thus not having any visual grounding and hence the similarity decreases. Finally, we observed a significant relationship between the levels of abstraction and the different tutor responses. However, the residuals did not reveal any further relations between the individual categories. We hypothesize that the abstraction is not the correct labeling for the dialogues in this case, as we could expect all the abstraction categories to correspond to each of the tutor responses as the students engage in the problem-solving. Instead, more nuanced measures may be needed to understand the student's process around addressing tutor response, such as what we found with the gaze data.

In this experiment, the students were working within an ITS environment. However, we believe that our results would also extend beyond ITSs to other technology enhanced environments that also collect log data. One limitation of the current methods is that using eye-tracking, as the technology currently stands, is not necessarily feasible outside of a lab setting where the use of the technology does not scale-up. For future work, we would like to be able to understand how the correlations from this process data also correspond to the student learning that occurs. To investigate this question, we would want to find the correlation between the combined process data events and student learning measures. The work presented in this paper contributes to the field through combining multiple process data streams to form a better understanding of student knowledge acquisition. By combining more discrete and continuous data sources that capture the collaborative interactions between both students and the system, we can better understand how these events interact to influence the student learning process.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. Journal of memory and language, 38(4), 419-439.

Belenky, D.M., Ringenberg, M., Olsen, J.K., Aleven, V., & Rummel, N. (2014). Using dual eye-tracking to evaluate students' collaboration with an Intelligent Tutoring System for elementary-level fractions. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Chi, M. T. & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, *49*(4), 219-243.

Cherubini, M., Nüssli, M. A., & Dillenbourg, P. (2008, March). Deixis and gaze in collaborative work at a distance (over a shared map): a computational model to detect misunderstandings. In *Proceedings of the 2008 symposium on Eye tracking research & applications* (pp. 173-180). ACM.

D'Angelo, S., & Begel, A. (2017). Improving Communication Between Pair Programmers Using Shared Gaze Awareness. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (pp. 6245-6290). ACM.

Diziol, D., Walker, E., Rummel, N., & Koedinger, K. R. (2010). Using intelligent tutor technology to implement adaptive support for student collaboration. *Educational Psychology Review*, *22*(1), 89-102.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological science*, *11*(4), 274-279.

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

Jaques, N., Conati, C., Harley, J. M., & Azevedo, R. (2014). Predicting affect from gaze data during interaction with an intelligent tutoring system. In International Conference on Intelligent Tutoring Systems (pp. 29-38). Springer, Cham.

Jermann, P., & Nüssli, M. A. (2012, February). Effects of sharing text selections on gaze cross-recurrence and interaction quality in a pair programming task. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work* (pp. 1125-1134). ACM.

King, A. (1999). Discourse patterns for mediating peer learning. In A.M. O'Donnell, A. King (Eds.), *Cognitive Perspectives on Peer Learning* (pp. 87–117). Lawrence Erlbaum Associates, Mahwah, NJ.

Meyer, A. S., Sleiderink, A. M., & Levelt, W. J. M. (1998). Viewing and naming objects: Eye movements during noun phrase production. *Cognition, 66*, B25-B33.

Murray, T. (2003). An overview of intelligent tutoring system authoring tools: Updated analysis of the state of the art. In T. Murray, S. B. Blessing, & S. Ainsworth (Eds.), Authoring tools for advanced technology learning environments (pp. 491-544). Springer Netherlands.

Olsen, J. K., Belenky, D. M., Aleven, A., & Rummel, N. (2014). Using an intelligent tutoring system to support collaborative as well as individual learning. In S. Trausan-Matu, K. E. Boyer, M. Crosby, & K. Panourgia (Eds), *Proceedings of the 12th International Conference on Intelligent Tutoring Systems,* (pp. 134-143). Berlin, Heidelberg: Springer.

Prieto, L. P., Sharma, K., & Dillenbourg, P. (2015). Studying teacher orchestration load in technology-enhanced classrooms. In *Design for Teaching and Learning in a Networked World* (pp. 268-281). Springer, Cham.

Raca, M., & Dillenbourg, P. (2013, April). System for assessing classroom attention. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 265-269). ACM.

Rau, M. A., Michaelis, J. E., & Fay, N. (2015). Connection making between multiple graphical representations: A multi-methods approach for domain-specific grounding of an intelligent tutoring system for chemistry. *Computers & Education*, *82*, 460-485.

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*(6), 1045-1060.

Richardson, Daniel C., Dale, R., & Kirkham, N. Z. (2007). The Art of Conversation Is Coordination. *Psychological Science, 18*(5), 407–413. doi:10.1111/j.1467-9280.2007.01914.x

Ritter, S., Anderson, J. R., Koedinger, K. R., & Corbett, A. (2007). Cognitive Tutor: Applied research in mathematics education. *Psychonomic bulletin & review*, *14*(2), 249-255.

Sangin, M., Molinari, G., Nüssli, M. A., & Dillenbourg, P. (2011). Facilitating peer knowledge modeling: Effects of a knowledge awareness tool on collaborative learning outcomes and processes. *Computers in Human Behavior*, *27*(3), 1059-1067.

Schneider, B., Sharma, K., Cuendet, S., Zufferey, G., Dillenbourg, P., & Pea, R. (2016). Using mobile eye-trackers to unpack the perceptual benefits of a tangible user interface for collaborative learning. ACM Transactions on Computer-Human Interaction (TOCHI), 23(6), 39.

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.

Sharma, K., Alavi, H. S., Jermann, P., & Dillenbourg, P. (2016). A gaze-based learning analytics model: in-video visual feedback to improve learner's attention in MOOCs. In Proceedings of the Sixth International Conference on Learning Analytics & Knowledge (pp. 417-421). ACM.

Sharma, K., Caballero, D., Verma, H., Jermann, P., & Dillenbourg, P. (2015a). Looking AT versus looking THROUGH: A dual eye-tracking study in MOOC context. In Proceedings of 11th International Conference of Computer Supported Collaborative Learning (Vol. 1, No. EPFL-CONF-203805, pp. 260-267). ISLS.

Sharma, K., Caballero, D., Verma, H., Jermann, P., & Dillenbourg, P. (2015b). Shaping learners' attention in Massive Open Online Courses. *Revue internationale des technologies en pédagogie universitaire/International Journal of Technologies in Higher Education*, *12*(1-2), 52-61.

Slavin, R. E. (1996). Research on cooperative learning and achievement: What we know, what we need to know. *Contemporary Educational Psychology, 21*(1), 43-69.

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, *46*(4), 197-221.