

CSCL and Eye-tracking: Experiences, Opportunities and Challenges

Kshitij Sharma, École Polytechnique Fédérale de Lausanne, kshitij.sharma@epfl.ch
Patrick Jermann, École Polytechnique Fédérale de Lausanne, patrick.jermann@epfl.ch
Pierre Dillenbourg, École Polytechnique Fédérale de Lausanne, pierre.dillenbourg@epfl.ch
Luis P. Prieto, School of Educational Sciences, Tallinn University (Estonia), lprisan@tlu.ee
Sarah D'Angelo, Northwestern University, sdangelo@u.northwestern.edu
Darren Gergle, Northwestern University, dgergle@northwestern.edu
Bertrand Schneider, Harvard Graduate School of Education, bertrand_schneider@gse.harvard.edu
Martina Rau, Educational Psychology, University of Wisconsin-Madison, marau@wisc.edu
Zach Pardos, Graduate School of Education, University of California-Berkeley, pardos@berkeley.edu

Nikol Rummel (discussant), Ruhr-Universität Bochum, nikol.rummel@rub.de

Abstract: The idea of using gaze as a medium to look into the collaborative processes had been around in CSCL for past few years. However, it had not been widely used in the community. Most of the works done in the direction of understanding collaborative cognition are majorly based on the qualitative methods. Research has shown that the collaborative gaze data can be used as an alternate source of information to assess collaboration. Once, we understand the how the gaze data reflects the collaboration quality and success, we could design gaze-aware systems to support remote/collocated collaboration. In this symposium, we bring together five papers that use eye-tracking data as a proxy for communication and cognition during remote/collocated collaborative learning and propose design of gaze-aware systems.

Keywords: eye-tracking, gaze-aware applications, modeling collaboration

Introduction

Eye-tracking has been used to explain the students' behavior in individual learning scenarios (Slykhuis et. al, 2005; Tsai et. al, 2012; Mayer, 2010). However, the use of eye-tracking in collaborative learning situations, more importantly in CSCL is rather uncommon. Findings of CSCL researchers have shown that eye-tracking can be equally useful to explain the collaborative learning processes as they are in individual learning cases.

Most of the methods employed to analyze and assess collaboration in CSCL are heavily based on qualitative methods. According to a review by Jeong and Hmelo-Silver (2010), 42% of the studies are based on qualitative analysis and/or "code and count" methods. These processes are tedious. Eye-tracking provides an automatic way of analyzing and assessing the collaboration, which could aid in gaining deeper and richer understanding of collaborative cognition. With the increasing number of eye-tracking studies, in collaborative settings, there is a need to create a shared body of knowledge about the relations found between gaze-based variables and cognitive constructs.

Eye-tracking has potential to be used in a wide subspace of CSCL "ecosystem", that is, different collaborative settings (remote or collocated); different learning instruments (multimedia, immersive or tangible); different learning formats (formal or informal); different planes (Dillenbourg et. al, 2011) in a classroom (individual, group, or the class); both sides of the instruction (teacher and/or student); tutoring systems and gaze-aware applications to support students in collaborative and/or individual settings. The use of eye-tracking will not only help us enriching our understanding about learning processes and relation of the gaze with the learning outcomes, but also proactively use students'/collaborators' gaze to inform them about their progress and mistakes.

In this symposium, the five papers cover the a large subset of CSCL "ecosystem", that is, remote and collocated collaborative settings, understanding teacher's orchestration, intelligent tutoring systems, and designing gaze-aware systems to support collaboration. The researchers will each present their work setting forth the conceptual, theoretical, practical advancements, and the challenges faced in eye-tracking research. The discussant will address how these papers have collectively advanced the chances of widening the use of eye-tracking in CSCL and analytics. The prime motivation of this symposium is to bring forward the findings, pitfalls, cautions, and challenges that appear while conducting eye-tracking research in CSCL and other formats of learning, to a wider range of audience, to present eye-tracking as an easy-to-go research tool; and finally, to burst the image of eye-trackers as a technology jargon.

Looking THROUGH versus looking AT

Kshitij Sharma, Patrick Jermann, Pierre Dillenbourg

Over multiple dual eye-tracking studies, researchers used dual eye-tracking as a proxy for cognition underlying collaboration. Richardson and Dale (2005), in a listening comprehension task, found that there was an eye-eye span (speaker's eye to listener's eye) of about 2 seconds. Jermann and Nüssli (2012) later confirmed this in a pair-programming task. There is a clear difference in the interaction styles of the teams that collaborate well in a given task and succeed; and the teams that do not and fail. In a collaborative program comprehension task Sharma et. al. (2012) found that the good pairs look at the data-flow of the program while the poor pairs read the program as if it was an English text. In the same study Sharma et. al. (2013) found that there is higher amount of similarity (probability to look at the same set of objects in the same time period) for the good pairs than that for the bad pairs. In a dual eye-tracking experiment, where participants first watched a video lecture individually and later collaboratively created a concept-map about the content of the lecture, we found that the participants who were individually following the teacher, both in deictic and dialogue space, more than the others, also had higher similarity during the collaborative concept-map task (Sharma et. al., 2015). These results indicate towards a common hypothesis that there exist two different ways of interacting with the content and the collaborator in a remote collaborative setting: "Looking THROUGH" and "Looking AT".

The concepts of "looking through" and "looking at" could be seen as new interaction style categories. "Looking at" the interface/display, indicates that the person is engaged with the material only, which is presented to him/her. "Looking through" the interface/display, indicates that the person is engaged with the peer. As an analogy, to highlight the difference between the two interaction styles, we can compare the interaction with the teacher/collaborating partner to watching a movie. "Looking at" can be compared with liking the movie; whereas, "looking through" can be compared with appreciating the director.

In a dual eye-tracking experiment with 120 students, as the first attempt to quantify these two interaction styles, we used two variables: "with-me-ness" (Sharma et. al., 2014) and "similarity" (Sharma et. al., 2013). The experimental task was to individually watch a video lecture first, and then to collaboratively create a concept map about the content of the video lecture.

With-me-ness is a measure for quantifying "how much are students following the teacher" during the video lectures. With-me-ness has two components: 1) perceptual with-me-ness and 2) conceptual with-me-ness. The perceptual with-me-ness captures the students' attention especially during the moments when the teacher makes explicit deictic gestures. Whereas, the conceptual with-me-ness captures whether and how much the student is following the teacher's dialogues. To compute conceptual with-me-ness, two authors mapped the teachers' dialogues to the different objects (objects of interest) on the screen. Once we have the objects of interest on the screen, we computed what proportion the dialogue length, (+2 seconds) in time, is spent by the participants on the objects of interest. This proportion is the measure of the conceptual with-me-ness.

Gaze similarity is the measure of how much the two participants in a pair were looking at the same thing at the same time or how similar their patterns were during a short period of time. To compute the similarity the whole interaction (during the collaborative concept map task) is divided into equal duration time windows. For each time window we compute a proportion vector, for each participant, containing the proportion of the window duration spent on each object of interest on the screen. Finally, the similarity is computed as the scalar product of the proportion vector for the two participants in a pair. Gaze similarity is a similar measure as the cross-recurrence proposed by Richardson and Dale (2005) but it is easier and faster to compute.

The results show a strong correlation between the average with-me-ness of the pair (during video lecture) and gaze similarity (during collaborative concept map). This suggests that there exist two categories of interaction styles: engaging with the material (looking at) or engaging with the peer (looking through). The peer in the video phase is the teacher and in the collaborative concept map is the collaborating partner. The "looking through" interaction resembles the social colocation of the interacting peers. A challenge for the next iteration of experiments would be to define semantic-less measures that could be applied to any context and not be restricted to video based instruction.

Gaze data on representational competencies in an intelligent tutor

Martina Rau, Zach Pardos

When students collaboratively solve problems in STEM, they often use visual representations (NRC, 2006). For example, students may collaboratively construct the visual representations shown in Figure 1 to solve chemistry problems. Students' difficulties with visual representations are well documented: they may not know how to interpret, construct, or reason with visual representations (for overviews, see [(Ainsworth, 2006). Thus, to benefit

from visual representations, students need *representational competencies* that enable them to learn with visual representations (Rau, 2016). Representational competencies are particularly important for collaborative learning. First, visual representations can enhance collaboration quality by allowing students to externalize their reasoning, which helps establish common ground (Suthers & Hundhausen, 2003). Second, collaboration can help students make sense of visual representations because divergent views can prompt deeper engagement in sense making of representations (Gnesdilow, Bopardikar, Sullivan, & Puntambekar, 2010).

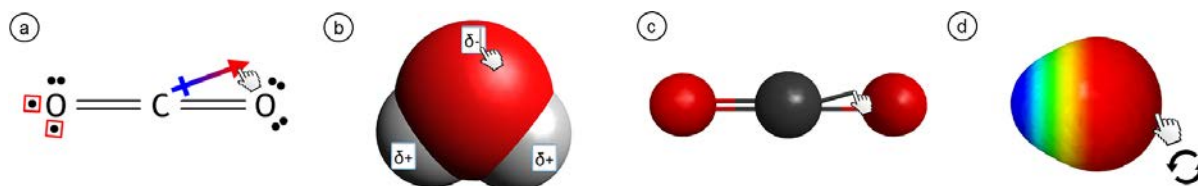


Figure 1. Visual representations in chemistry: a: Lewis structure; b: space-filling model; c: ball-and-stick model, d: EPM.

Intelligent tutoring systems (ITSs) offer several functionalities that can support collaborative problem solving, such as adaptive collaboration scripts (Walker, Rummel, & Koedinger, 2009) and just-in-time help and feedback (VanLehn, 2011). A new trend in research on ITSs is to support students' representational competencies (Rau, in press). ITSs adapt to students' needs for support based on a cognitive model that infers each student's knowledge level based on his/her interactions within the system (VanLehn, 2011). A limitation of current cognitive models is that they capture students' domain knowledge, not on representational competencies. Hence, current ITSs can adapt to domain knowledge but not to representational competencies. Yet, in light of the key role of representational competencies in STEM learning, providing support that adapts to students' representational competencies may significantly enhance the effectiveness of ITSs.

It seems reasonable to assume that we can gather useful information about students' representation competencies from their visual attention to representations. Prior eye-tracking research has several limitations that leave this question open. First, most prior eye-tracking research involved relatively simple learning materials (e.g., expository text paired with a static visual representation; [Mason, Pluchino, Tornatora, & Ariasi, 2013; Mayer, 2010]. By contrast, ITSs are more complex because they involve multiple interactive visual representations (see Figure 1). Second, prior eye-tracking research shows that eye- data improves the accuracy of cognitive models that assess students' domain knowledge [e.g., (Bondareva et al., 2013). Yet, cognitive models can use students' interactions to assess their representational competencies. It has not been tested whether eye- data can improve the accuracy of a cognitive model of representational competencies.

We found first indications that data might predict students' learning came in a study in which 25 undergraduate chemistry students worked with an ITS for chemistry. The ITS contained a cognitive model that captured students' representational competencies. Students worked with the ITS for 2.5h while a SMI RED 250 collected eye- data. Results showed that durations of students' fixations on visual representations predicted error rates obtained from the ITS log data, which in turn predicted students' learning outcomes on a domain-knowledge posttest. In a second study, we tested whether adding data to the ITS's cognitive model would enhance the model's accuracy in predicting students' errors when they solved problems within the ITS. 95 undergraduate students worked with the ITS for 3h. Results revealed no added benefit of adding eye- features (e.g., frequency of switching between representations, fixation durations in specific representations) to the cognitive model in terms of accuracy in predicting students' errors during problem solving. A limitation of this study was that the eye- features were available only at the level of the ITSs' problem, which contained multiple steps. Hence, in a third study, we tested whether more fine-grained eye- features at the level of problem-solving steps would improve the cognitive model's accuracy. 117 undergraduate students worked with the ITS for 3h. Results showed no added benefit of adding eye- data to the cognitive model.

Taken together, these findings stand in contrast to prior research that has found that eye- data can enhance the accuracy of cognitive models [e.g., (Bondareva et al., 2013). One important difference to our research is that our cognitive model assessed students' representational competencies. Hence, our findings may suggest that adding eye- data to a cognitive model that captures representational competencies based on students' interactions with visual representations. This rationale amounts to a new hypothesis that should be tested in future research: namely that adding representational competencies to a cognitive model of domain knowledge may improve the model's accuracy as much as the addition of eye- data would.

Dual eye-tracking in co-located spaces

Most dual eye-tracking studies remote collaborations involve two participants looking at two computer screens. This setup has a relatively low ecological validity, because most collaborative tasks still happen in co-located settings (e.g., face-to-face or side by side). Thus, it is difficult to know whether the results from remote collaborations actually generalize to co-located interactions. This gap in the literature is mostly the result of technical challenges: researchers can easily know whether two participants are looking at the same things on a screen, because the computer has perfect knowledge of what is displayed. In the real world, however, the computer has no knowledge of what is being captured by the camera of a mobile eye-tracker. Consequently, it is much more difficult to tell whether two participants are actually looking at the same location or not. In our own work (Schneider, Sharma, Cuendet, Zufferey, Dillenbourg & Pea, accepted), we have conducted an empirical study where apprentices in logistics ($N=54$) interacted with a Tangible User Interface (TUI). By leveraging the fiducial markers used by the TUI, we were able to remap students' s onto a ground truth (bottom half of Fig. 2, left side). This allowed us to replicate the results found in remote collaborations. We found that groups who had higher levels of joint visual attention tended to have a higher quality of collaboration, do better at the task given to them, and learn more from it. Additional results also suggested that students who used a 2D version of the TUI (i.e., with flat paper shelves instead of 3D shelves) tended to have less moments of JVA compared to students who used a 3D version of the tangible interface. While we are still investigating this effect, this result has interesting implications for designing collaborative interfaces: if one's goal is to support visual coordination in groups of students or collaborators, our findings suggest that 2D interfaces (such as computer screens, tablets) may not have the same affordances as environments that exhibit some 3D structure. Instead, 3D physical objects or environment might be best suited for collaborative work.

Dual eye-tracking datasets also allow researchers to identify particular group dynamics (Schneider, Sharma, Cuendet, Zufferey & Dillenbourg, 2016). Using the same dataset, we were able to first develop an enhanced version of a cross-recurrence graph. Two groups are contrasted on Figure 2 (right side): Both had high levels of JVA, but the group on the left had below average learning gains while the group on the right had above average learning gains. The top row shows the traditional cross-recurrence graph, while the middle rows shows our augmented version. Colors show where students had a moment of JVA (red means that participants were jointly looking at the leftmost warehouse on Fig. 2 (left side), green is for the warehouse in the middle and blue for the rightmost warehouse). Dotted squares indicate when the experimenter provided students with prompts. Color-coding offers insights about the strategies used by students: group 13 spent a lot of time going back and forth between warehouses, while group 20 focused on one model at the time. Since the goal was to identify design principles by *comparing* those layouts, it is not surprising that group 13 did better on the learning test. Finally, the bottom row of Fig. 2 (right side) shows speech data from each participant.

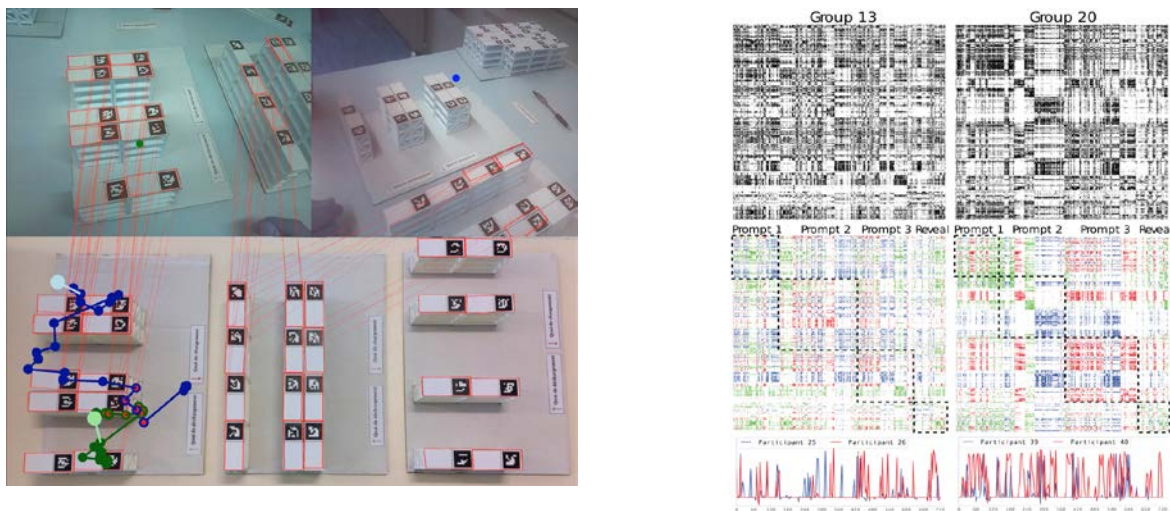


Figure 2. On the left: remapping two s onto a ground truth using two synchronized mobile eye-trackers. The top left image is the perspective of the first student, the top right image is the perspective of the second student, and the bottom image is the ground truth. Red dots show joint visual attention, and line between the three perspectives show common points used to remap students' s onto the ground truth. On the right: traditional cross-recurrence graphs (top), augmented with spatial information (red, green, blue) and speech (bottom).

In both groups, one participant (in red) tended to talk more while the other person (in blue) was quieter. By looking at the transcript, we realized that this pattern hid some crucial differences between those two groups. While the blue participant in group 20 would always agree with his partner, the blue participant in group 13 would constantly challenge his partner by pointing at counter-examples. So in one case, there was a clear free-rider effect where the more passive participant was intellectually disengaged from the activity. In the other group, the more passive student was actually actively contributing. We found that this pattern could be found in the eye-tracking data: for each moment of JVA, we identified who initiated it (i.e., whose was there first) and who responded to it (i.e., whose was there second). We found a significant correlation between learning gains and students' tendency to equally share the responsibility of initiating and responding to offers of JVA. In other words, groups where the same person always initiated moments of JVA were less likely to learn (e.g., group 20) and groups where this responsibility was evenly shared were more likely to learn (e.g., group 13). This finding shows that we can go beyond merely quantifying JVA, and actually identify (counter-) productive group dynamics using dual eye-tracking data in co-located settings.

In summary, there are some new interesting efforts pushing the boundaries of what has been previously done in the study of JVA. The first generation of studies was qualitative by nature, and used time-consuming analyses of videos to provide a detailed account of the micro-genesis of JVA (most notably with babies). The second generation started to use synchronized eye-trackers to quantitatively describes visual coordination and provide correlates of collaboration quality. We currently seeing a third wave of studies using synchronized eye-trackers, where those sensors are used to design interventions to support social interactions and where mobile eye-trackers are used to quantify JVA in co-located settings. Those new developments open new exciting doors to both capture and influence JVA in a variety of settings.

Designing representations for remote learning

Sarah D'Angelo and Darren Gergle

Integrating *awareness* into remote learning environments is one way to introduce missing non-verbal cues that are leveraged in effective co-located learning. This technique involves collecting eye movement data from people working on the same task and visually representing that information on screen for collaborators. Sharing patterns may be particularly helpful when remote teachers are explaining linguistically complex visual elements where it is difficult to create shared understanding. In dyadic interactions the basis of shared understanding is often the development of common ground among a pair. Explicit deictic gestures or references (e.g. pointing and saying "here") play a key role in establishing and maintaining common ground (Clark & Brennan, 1991). Therefore, a way to help students understand deictic references in remote environments is to display where the teacher is looking, because information can help the listeners better disambiguate deictic references in complex visual environments (Gergle & Clark, 2011; Hanna & Brennan, 2007). In this work, we explore displaying the teacher's information as a video augmentation to aid in understanding complex visual content and to help students follow along with the teacher, maintain attention, and model approaches used by experts when examining visual content.

We designed a video lecture on cloud identification that allowed us to evaluate the utility of video augmentations with highly visual and linguistically complex content. We evaluate two deixis visualizations (pointer and) in the context of a MOOC style video lecture on visually complex content (cloud identification). Deixis visualization is a representation of a physical gesturing (pointing with a pen) or a shift in attention (looking in a specific place) that is coupled with a deictic reference (e.g. "here"). The results suggest that showing the teacher's to students when making explicit references to information on the slides can be useful for students. When shown the teacher's information, students scored higher on the posttest compared to no visual aid. Additionally, students in the condition spent more time looking at relevant points and had similar patterns to the teacher. This suggests that the visualization helped students follow along with the lecture and helped them to look at the visual information in the appropriate way.

One possible explanation for this improvement in performance is the design of the visualization. We displayed the teacher's scan path, which highlights more areas on the slides by illustrating where the teacher was previously looking which would not be highlighted using pointer representation. This additional information may have helped students connect and integrate relevant information. In comparison, a recent study used a single point representation in a real-time collaborative learning exercise between peers. Their results show that awareness increased the partners' joint attention and improved gains in learning compared to no display (Schneider & Pea, 2013). This suggests that the design of visualizations can support the specific learning task. In a teacher to student learning task, a scan path representation helped students follow along with the teacher and learn appropriate eye movement patterns. On the other hand, in the collaborative learning study between peers, the single point

representation helped students communicated about complex concepts. This raises the question; *how should representations be designed to effectively support different types of learning tasks?*

In this presentation, I will draw on results from my work and others as well as future directions to discuss the importance of creating effective visualizations of to support remote collaborative learning. For example, the degree of coupling between the pair might determine if continuous awareness is harmful or helpful (Brennan, Chen, Dickinson, Neider, & Zelinsky, 2008; D'Angelo & Gergle, 2016). Additionally, timing and distribution of knowledge can alter how pairs interpret another person's, which can illustrate a trajectory over time or a real time point of attention (Schneider & Pea, 2013; Stein & Brennan, 2004). Therefore, it is important to consider the features of the learning task when designing visual representations of to support effective collaboration between teachers and students.

Eye-tracking in CSCL orchestration research: Raw metrics and automation potential

Luis P. Prieto, Kshitij Sharma, Pierre Dillenbourg

Although CSCL puts most emphasis on learners and their interactions as crucial elements in learning, multiple studies (especially, in formal education settings) have highlighted the crucial role of teacher orchestration in its effectiveness (e.g., Onrubia & Engel, 2012). This has led to an interest in research about teacher orchestration of CSCL, both as an essential issue in the design of CSCL and other educational technologies (Dillenbourg et al., 2011) as well as in the process of implementing and evaluating CSCL innovations, taking into account the multiple constraints of everyday educational practice in authentic educational settings (Roschelle, Dimitriadis, & Hoppe, 2013).

Eye-tracking has been traditionally linked to lab studies in controlled conditions; recent advances in mobile eye-trackers (often in the form of wearable goggles) allow the study of eye movements in more natural settings. Indeed, such technologies are already being used in fields like human-computer interaction (HCI), usability engineering or marketing studies. The rest of this section describes two strands of research in which we have used such mobile eye-trackers to study the orchestration of learning (very often, CSCL) processes 'in the wild' (i.e., in authentic classroom settings).



Figure 3. Classroom setup in one of the classroom studies to assess teacher's orchestration load through a mobile wearable eye-tracker (left, in the center of the image). Planned, human-coded and automatically coded orchestration graph of a collaborative learning lesson (right).

One of the main emphases of orchestration-related CSCL research is the technology design perspective: aside from being good for learners' outcomes, collaborative learning technologies that intend to be applied in everyday educational practice, have to be usable "at the classroom level" (Dillenbourg et al., 2011). That is, they also have to comply with multiple other restrictions of classroom settings, such as time or discipline constraints, or the limited cognitive resources of a teacher trying to keep track and support multiple collaborative learning processes at the same time (what some authors call 'orchestration load'). However, so far the output of this kind of research has been mainly in the form of abstract, high-level design principles distilled from the observation of isolated classroom experiments. The increasing availability of mobile eye-trackers, plus the work in psychology and HCI relating cognitive load and eye-related measures has the potential to provide a quantitative measure of orchestration load in real classroom environments. In our own work, we have explored the use of mobile eye-trackers to estimate teachers' orchestration load in diverse authentic classroom settings, from collaborative tabletop games to more usual laptop-based classrooms (e.g., Luis P. Prieto, Sharma, Wen, & Dillenbourg, 2015) (see Figure 3, left). Contrary to other contributions in this symposium, in this work we do not exploit so much

what the subjects are looking at, but rather the raw, low-level physiological measures of *how* the teacher looks at students and classroom elements, which can provide insight into cognitive or even social aspects of teacher-student interactions in the orchestration of collaboration.

The aforementioned studies use (manual) video coding analyses to give context to the low-level raw measures of pupil and eye movement, and to add a semantic layer to the trends in these measures (e.g., what kind of episodes have tendency to be high-load or low-load). For instance, one of the noticeable trends in most of our studies is that class-level interactions tended to be high-load (as opposed to interactions with small groups of students). However interesting these results are for our understanding of the challenges in orchestrating CSCL processes, this manual coding makes it difficult to gather data at scale (e.g., in order to assess the generalizability of results) or apply the approach to the everyday practice of our schools. Recent advances in wearable and ubiquitous sensors, machine learning and computer vision, however, may help us to *automatically* infer certain aspects of the social and behavioral context of orchestration actions directly from the eye-tracker/sensor data. We have explored the feasibility of this approach to automate the coding of teaching activities and social planes of interaction (see Figure 3, right), obtaining reasonable levels of accuracy even with relatively simple algorithms and feature extraction (L.P. Prieto, Sharma, Dillenbourg, & Rodríguez-Triana, 2016). This emergent path of research (what we call ‘multimodal teaching analytics’) illustrates the potential of eye-tracking data for semi-automated analysis. Future advances in this kind of automation may in turn improve the scalability and reach of CSCL research efforts in authentic classroom conditions, both to improve our understanding of teacher decision-making and experience in the orchestration of CSCL processes (an under-developed area of research), and have direct applications for teacher education and professional development (e.g., through evidence-based reflection based on eye-tracking data gathered from everyday practice).

References

- Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction*, 16(3), 183-198.
- Bondareva, D., Conati, C., Feyzi-Behnagh, R., Harley, J. M., Azevedo, R., & Bouchet, F. (2013). Inferring learning from data during interaction with an environment to support self-regulated learning. In H. C. Lane, K. Yacef, J. Mostow & P. Pavlik (Eds.), *Artificial Intelligence in Education* (pp. 229-238). Berlin Heidelberg: Springer.
- Brennan, S. E., Chen, X., Dickinson, C. A., Neider, M. B., & Zelinsky, G. J. (2008). Coordinating cognition: The costs and benefits of shared during collaborative search. *Cognition*, 106(3), 1465-1477.
- Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. *Perspectives on Socially Shared Cognition*, 13(1991), 127-149.
- D’Angelo, S., & Gergle, D. (2016). d and Confused: Understanding and Designing Shared for Remote Collaboration. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 2492-2496). ACM.
- Dillenbourg, P., Zufferey, G., Alavi, H., Jermann, P., Do-Lenh, S., Bonnard, Q., ... Kaplan, F. (2011). Classroom orchestration: The third circle of usability. In *Proceedings of the International Conference of Computer-Supported Collaborative Learning (CSCL2011)*.
- Gergle, D., & Clark, A. T. (2011). See What I’m Saying?: Using Dyadic Mobile Eye Tracking to Study Collaborative Reference. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work* (pp. 435-444). New York, NY, USA: ACM.
- Gnesdilow, D., Bopardikar, A., Sullivan, S. A., & Puntambekar, S. (2010). *Exploring convergence of science ideas through collaborative concept mapping*. Paper presented at the 9th International Conference of the Learning Sciences.
- Hanna, J. E., & Brennan, S. E. (2007). Speakers’ eye disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language*, 57(4), 596-615.
- Jeong, H., & Hmelo-Silver, C. E. (2010, June). An overview of CSCL methodologies. In *Proceedings of the 9th International Conference of the Learning Sciences-Volume 1* (pp. 921-928). International Society of the Learning Sciences.
- Jermann, P., and Nüssli, M.-A. (2012). Effects of sharing text selections on cross-recurrence and interaction quality in a pair programming task. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 1125-1134. ACM.
- Mason, L., Pluchino, P., Tornatora, M. C., & Ariasi, N. (2013). An eye-tracking study of learning from science text with concrete and abstract illustrations. *The Journal of Experimental Education*, 81(3), 356-384.
- Mayer, R. E. (2010). Unique contributions of eye-tracking research to the study of learning with graphics. *Learning and Instruction*, 20(2), 167-171.

- NRC. (2006). *Learning to Think Spatially*. Washington, D.C.: National Academies Press.
- Onrubia, J., & Engel, A. (2012). The role of teacher assistance on the effects of a macro-script in collaborative writing tasks. *International Journal of Computer-Supported Collaborative Learning*, 7(1), 161–186.
- Prieto, L. P., Sharma, K., Dillenbourg, P., & Rodríguez-Triana, M. J. (2016). Teaching analytics: Towards automatic extraction of orchestration graphs using wearable sensors. In *ACM International Conference Proceeding Series* (Vol. 25–29–Apri). <http://doi.org/10.1145/2883851.2883927>
- Prieto, L. P., Sharma, K., Wen, Y., & Dillenbourg, P. (2015). The burden of facilitating collaboration: towards estimation of teacher orchestration load using eye-tracking measures. In *Proceedings of the 11th international conference on computer-supported collaborative learning (CSCL 2015)* (pp. 212–219).
- Rau, M. A. (2016). Conditions for the effectiveness of multiple visual representations in enhancing STEM learning. *Educational Psychology Review*, 1–45. doi: 10.1007/s10648-016-9365-3
- Rau, M. A. (in press). A framework for discipline-specific grounding of educational technologies with multiple visual representations. *IEEE Transactions on Learning Technologies*.
- Richardson, D. C., and 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.
- Roschelle, J., Dimitriadis, Y., & Hoppe, U. (2013). Classroom orchestration: Synthesis. *Computers & Education*, 69, 523–526. article. <http://doi.org/http://dx.doi.org/10.1016/j.compedu.2013.04.010>
- Schneider, B., & Pea, R. (2013). Real-time mutual perception enhances collaborative learning and collaboration quality. *International Journal of Computer-Supported Collaborative Learning*, 8(4), 375–397.
- Schneider, B., Sharma, K., Cuendet, S., Zufferey, G., Dillenbourg, P., & Pea, R. (2015). Detecting Collaborative Dynamics Using Mobile Eye-Trackers. In *Proceedings of the 12th International Conference of the Learning Sciences*.
- Schneider, B., Sharma, K., Cuendet, S., Zufferey, G., Dillenbourg, P., & Pea, R. (accepted). Unpacking The Perceptual Benefits of a Tangible Interface. *ACM Transactions on Computer-Human Interactions*.
- Sharma, K., Jermann, P., Nüssli, M. A., & Dillenbourg, P. (2012). Evidence for different activities in program understanding. In 24th Annual conference of Psychology of Programming Interest.
- Sharma, K., Jermann, P., Nüssli, M. A., & Dillenbourg, P. (2013). Understanding collaborative program comprehension: Interlacing and dialogues. In *Computer Supported Collaborative Learning (CSCL 2013)*.
- Sharma, K., Jermann, P., & Dillenbourg, P. (2014). “With-me-ness”: A -measure for students' attention in MOOCs. In *International conference of the learning sciences* (No. EPFL-CONF-201918).
- Sharma, K., Caballero, D., Verma, H., Jermann, P., & Dillenbourg, P. (2015). Looking AT versus Looking THROUGH: A Dual Eye-Tracking Study in MOOC Context. In *Proceedings of 11th International Conference of Computer Supported Collaborative Learning, Gothenburg, Sweden, CSCL*.
- Slykhuis, D. A., Wiebe, E. N., & Annetta, L. A. (2005). Eye-tracking students' attention to PowerPoint photographs in a science education setting. *Journal of Science Education and Technology*, 14(5-6), 509-520.
- Stein, R., & Brennan, S. E. (2004). Another person's eye as a cue in solving programming problems. In *Proceedings of the 6th international conference on Multimodal interfaces* (pp. 9–15). ACM.
- Suthers, D. D., & Hundhausen, C. D. (2003). An experimental study of the effects of representational guidance on collaborative learning processes. *Journal of the Learning Sciences*, 12(2), 183-218.
- Tsai, M. J., Hou, H. T., Lai, M. L., Liu, W. Y., & Yang, F. Y. (2012). Visual attention for solving multiple-choice science problem: An eye-tracking analysis. *Computers & Education*, 58(1), 375-385.
- VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems and other tutoring systems. *Educational Psychologist*, 46(4), 197-221.
- Walker, E., Rummel, N., & Koedinger, K. R. (2009). CTRL: A research framework for providing adaptive collaborative learning support. *User Modeling and User-Adapted Interaction*, 19(5), 387-431.