International Society of the Learning Sciences

Challenging Joint Visual Attention as a Proxy for Collaborative Performance

Kshitij Sharma, Norwegian University of Science and Technology, kshitij.sharma@ntnu.no
Jennifer K. Olsen, University of San Diego, jenniferolsen@sandiego.edu
Himanshu Verma, Delft University of Technology h.verma@tudelft.nl
Daniela Caballero, University of Chile, daniela.caballero@ciae.uchile.cl
Patrick Jermann, École Polytechnique Fédérale de Lausanne, patrick.jermann@epfl.ch

Abstract: Researchers have used joint visual attention (JVA) as a proxy for collaborative quality and/or performance during the last decade due to its association with both measures. However, the notion of looking at the same object does not necessarily indicate that students are solving the problem together (or learning together). We propose a complementary approach to joint visual attention by augmenting it with joint mental effort (JME). JME is computed as a cross-recurrence of the cognitive load of the peers in a dyad. We use data from 41 dyads to show the synergy between JVA and JME and the insights that they can shed in the collaborative process. The results show that in certain episodes of collaboration (characterized by the dialogue and division of labor strategy of the dyad) combining these two dual-eye tracking measures provide deeper insights about the collaborative processes and performance than JVA alone.

Keywords: Dual eye-tracking, Collaborative Processes, Joint Visual Attention, Cognitive load

Introduction

It is not a new idea that collaborative learning can be beneficial for student learning outcomes nor within the computer-supported collaborative learning (CSCL) community that technology can be used to support the collaborative process (Johnson & Johnson, 1987; Lou, Abrami, & d'Apollonia, 2001). However, it continues to be an on-going process to understand what a productive collaboration process entails. As technology and computational methods continue to develop, our ability to measure the collaborative learning process changes by measuring it through different modalities and over time (Olsen, Sharma, Rummel, & Aleven, 2020; Starr, Reilly, & Schneider, 2018). A common motivator for finding new means of assessing the quality of the collaboration process has been the difficulty and time-consuming process of analysing student dialogue, especially in real time to be able to put interventions in place (Sharma et al., 2017). In these cases, the new measures are often a proxy for the dialogue content. When we consider collaboration measures in this way, we either intentionally or unintentionally assign moments of silence as less valuable. Rather, new measures of collaboration that can complement existing measures can fill in these gaps. In this paper, we aim to deepen our understanding of what effective collaboration looks like through the assessment of dual eye tracking measures. Furthermore, by assessing these eye tracking measures, we contribute to the understanding of how eye tracking can be used to analyse the collaborative learning process.

One collaborative learning theory is that students are able to assess and update their mental models of the domain by working with peers (Chi & Wylie, 2014). Through the process of co-construction, students can reflect on their own mental model to make repairs, incorporate their partner's ideas into their model, and construct new knowledge by building upon their partner's ideas (Hausmann, Chi, & Roy, 2004). In this case, the benefits of collaboration come from the joint construction of knowledge that occurs as students work together. To measure these processes, we can analyse episodes of interaction for indicators of students integrating their partner's ideas into their thought process. For example, one can measure collaboration through the use of transactivity (Joshi & Rosé, 2007) or through interactive dialogue as proposed in the ICAP framework (Chi & Wylie, 2014). Many of these coding schemes focus on the different ways in which students can construct knowledge. In a collaborative setting, this is just one aspect of the collaboration with students also needing to coordinate the work and in CSCL settings, coordinate with the technology (Rummel et al., 2011). Across all of these aspects, researchers have mainly used dialogue content as the gold star measure for collaboration but interactions with the activity are also common. When other collaborative learning measures are proposed, too often they are used as a proxy for analysing the dialogue content rather than complementing it (Sharma et al., 2017). The work in multi-modal learning analytics begins to address this gap by investigating how multiple modalities of data can be used together. However, before combining data streams, it is important to understand what each data stream can provide.

In this paper, we focus on the information that can be provided through dual eye tracking (DUET). In previous research, DUET has been used as a tool to explain the socio-cognitive mechanisms underlying collaborative learning (Jermann and Nuessli, 2012; Sharma et al, 2018 & 2020; Olsen et al, 2018). Information extracted from DUET data has been used to explain collaboration quality (Schneider et al, 2019), collaborative task-performance (Sangin et al., 2011; Jermann and Nuessli, 2012; Sharma et al, 2017), and collaborative learning gains (Olsen et al., 2020). Duet also has been used to explain certain processes related to collaborative learning, for example, mutual modelling (Lemaignan and Dillenbourg, 2015), repairs of misunderstanding (Cherubini et al., 2008), shared understanding (Richardson et al., 2007), and coordination (Brennan, et al., 2008). Additionally, researchers have used DUET as a method to provide collaborative awareness to the peers attempting to solve a given problem (Schneider and Pea, 2015, D'Angelo and Begel, 2017; D'Angelo and Gergle, 2016). In most of these studies, the basic outcome or the working hypothesis is that Joint Visual Attention (JVA) is a decent proxy of collaborative mechanisms. All of these studies emphasize a social extension of the eye-mind hypothesis, "what you see is what you process", to "looking together is processing together". However, this notion has not been verified in some studies over the past few years (e.g., Belenky et al, 2014). In this contribution, we revisit the concept of JVA and complement it with another DUET measurement, Joint Mental Effort (JME). This measurement is inspired by the Kirscher's view (Kirschner et al., 2018) of how transactive activities can exert cognitive loads on collaborating peers and that the absence of synchrony in the collaboration can be detrimental for collaborative performance (Popov et al., 2017). JME provides an attempt to create a proxy for the collaborative cognitive load synchrony.

Specifically, in this paper, we investigate what JVA and JME, both collected through eye tracking, indicate about the collaborative learning process. We analysed 82 master students working in pairs to construct a concept map related to the resting membrane potential. We were interested in how their collaborative process impacted the quality of their concept map. To measure the collaborative process, we collected student dialogues, eye tracking data and computer logs. In this paper, we aimed to answer two research questions through our analysis. First, how did our eye tracking measures (JVA and JME) relate to student performance? Second, how do JVA and JME relate to other indicators of collaboration, such as dialogue content and division of labor, and how do the interactions with student performance associate with JVA and JME? Based on previous studies, we hypothesize that JVA will be positively related to student performance (Richardson et al., 2007; Jermann and Nuessli, 2012) and that JME also will be positively associated with performance (Kirschner et al, 2018; Popov et al, 2017). Based on the results of these research questions, we discuss the benefits of using eye tracking measures to assess the collaborative learning process.

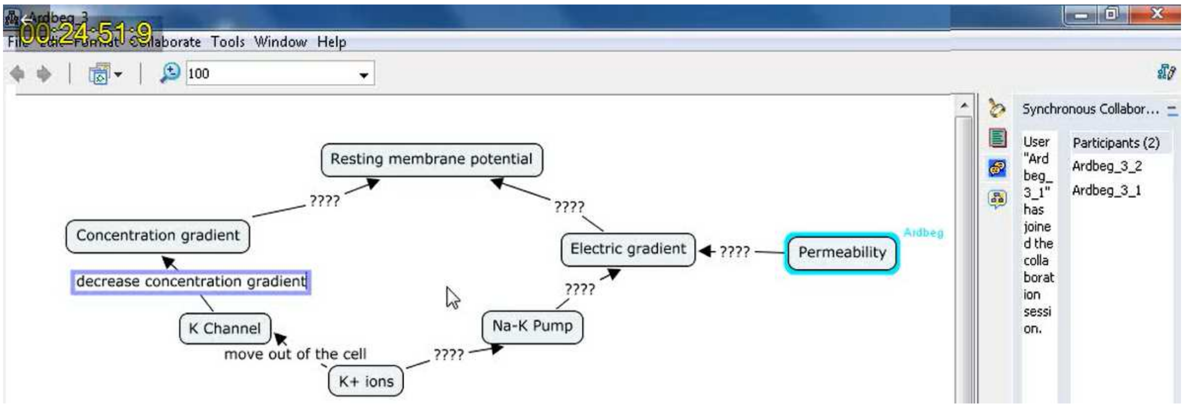
Methodology

Participants and procedure

We had 82 master students from École Polytechnique Fédérale de Lausanne participating in the present study in pairs. Of these students, 16 were female. Before beginning to create the concept map collaboratively, all participants individually watched two videos about "resting membrane potential", a topic about which the students did not know prior to participating in the task. Each video was 17.05 minutes long and provided the students with the information they would need for the development of the concept maps. While watching the videos, the participants had full control over the video player and no time constraints. After both partners completed the videos, they were asked to create a collaborative concept-map using IHMC CMap tools (Figure 1, top). The collaborative concept-map phase was 10-12 minutes long. Although each student remained working at their own computer, the participants could talk to each other while their screens were synchronized, i.e., the participants in the pair were able to see their partners' actions. There were 14 concepts preloaded on the Concept map tool and the main task for the pairs was to connect the given concepts with correct relationships. They could also add new concepts if they wanted.

Data collection

From the interaction of the dyad with the concept map tool, we collected the following data. 1) We collected eye-tracking data using two SMI remote eye-tracking devices (SMI RED 250) at the sampling rate of 250 Hz. For each participant, we use a 5-point calibration and a 5-point validation mechanism. The fixation and saccades were identified using the built-in algorithm of the BeGaze software. 2)We recorded the audio of the students' dialogues using the system audio from one of the computers. 3) We recorded all the actions done by the dyad on the concept map. The logs contained the timestamp of the action, peer ID, action type (add, delete, move, resize, text edit), conceptID and metadata (Figure 1, bottom). For example, if a student adds two concepts with a link. The system would log the time the action took place, the ID of the students, the action is logged as an "add" action, the object will be the "connection", there will be a new ID generated for this connection, and the metadata would log the two concepts it linked.



timestamp	SubjectID	Action	Object	ObjectID	Metadata
Mon Oct 28 19:11:30:168	Clynelish_1_1	Move	Concept	ge:1M5NG8B0L-1BJ5XH8-B9	'Resting membrane potential' x:134 y:43 w:188 h:28
Mon Oct 28 19:11:52:232	Clynelish_1_2	Add	Connection	ge:1M88V4FYD-3G09QV-PT	from:ge:1M88V4B39-1JK1H3M-KR to:ge:1M88V4FY8-G83G3M
Mon Oct 28 19:11:53:272	Clynelish_1_2	Move	Concept	ge:1M88V4FY8-G83G3M-PH	'????' x:516 y:280 w:44 h:26
Mon Oct 28 19:12:23:356	Clynelish_1_1	Add	Concept	ge:1M88V5H7G-MLDB4J-2TH	'????' x:488 y:271 w:44 h:26
Mon Oct 28 19:12:24:751	Clynelish_1_1	Delete	Concept	ge:1M88V5H7G-MLDB4J-2TH	'????' x:488 y:271 w:44 h:26
Mon Oct 28 19:12:31:756	Clynelish_1_1	Modify Text	Linking Phrase	ge:1M88V4B39-1JK1H3M-KR	'' x:512 y:282 w:4 h:18
Mon Oct 28 19:12:31:845	Clynelish_1_1	Move	Concept	ge:1M5NG9QK9-Y0RB8H-FX	'Cl channel' x:469 y:325 w:80 h:26

Figure 1. (Top) An example of the concept map under construction in the CMap tools. The two participants' names are on the top-right side, and their pointers have different colors. Whenever they perform an action, the relevant object (concept or link) is highlighted. (Bottom) Snapshot of the log file produced from CMap.

Measurements

All the measurements were computed at the dyad level, the time-unit for each computation was one utterance, and all the measurements were aggregated for the dyad.

Cognitive load similarity (CLS): Joint mental effort (JME)

From the eye tracking data, we calculated the students' JME, *a measure of the cognitive load similarity*. To calculate this measure, we first compute the individual cognitive load from the pupil dilation data using the method found in Duchowski et al. (2018). Next, we discretize the value to represent an integer value in the range zero to ten. Once we have the cognitive load for both peers in the dyad, we compute the cross-recurrence between the two time-series, using the method proposed by (Richardson et al., 2007).

Gaze similarity (GS): Joint visual attention (JVA)

JVA is a measure of *how similar two individual gaze patterns are.* In order to compute the similarity between the gaze patterns of two collaborating students, we computed the similarity between the two proportionality vectors discussed above by using the reverse function (1/(1+x)) of the correlation matrix of the two vectors (where x is the distance between the two proportionality vectors). A similarity value of 1 shows complete similarity between the two gaze patterns during a given time window. A lower value of similarity shows that the two participants spent less time looking at a similar set of objects on the screen during the same time window.

Dialogue codes

One of the authors transcribed the audio data and two authors coded the dialogues. The intercoder-consistency between the two coders was 0.86 (for 20% dyads). The dialogues were coded based on the fact whether the dyad is talking about the concept map tool and aesthetics (CMAP) or about the content of the concept map (KNWL). For example, *"Let's write something to remove the question marks"* would be coded as CMAP, and *"Resting membrane potential is the equilibrium between Na+ and Cl-" would* be coded as KNWL.

Division of labor (DoL)

Following a definition provided by Jermann (2004), we compute the division of labor using the number actions taken on a specific concept by one member of the dyad. Specifically, we compute the Sum of Differences (SD) and Sum of absolute differences (SAD) between members of a dyad using the formulae (1) and (2). Using thresholds on SD and SAD, we define three DoL levels, role, task and concurrent, which we outline in more detail below.

$$SD = \frac{\sum_i (S_1 C_i - S_2 C_i)}{S_1 C + S_2 C} \ (1) \quad SAD = \frac{\sum_i |S_1 C_i - S_2 C_i|}{S_1 C + S_2 C} \ (2)$$

In formulae (1) and (2), S1 and S2 are the peers in a dyad. C is the concept. S1C and S2C are the total number of actions done by peers S1 and S2, respectively. S1Ci and S2Ci are the actions done on concept Ci by S1 and S2, respectively. SD has a range of [-1, +1] with -1 indicating that S2 did all the actions, +1 indicating that S1 did all the actions and 0 depicting equal participation. SAD has a range of [0, 1] with 0 indicating equal participation and 1 indicating that all the actions were done by one peer.

We defined the DoL strategies – role, task and concurrent – based on SD and SAD values. The DoL strategy is classified as *role* if SAD is in the range [0.5, 1] and SD in either [0.33, 1] or [-1, -0.33] indicating that one student did all of the actions within a certain time window - implying the other student was either a free-rider or acting as a navigator. The DoL strategy is classified as *concurrent* if SAD is in the range [0, 0.5] and SD in range [-0.33, 0.33] during the time window, indicating that the students had equal participation on the same concepts. Finally, the DoL strategy is classified as *task* if SAD is in the range [0.5, 1] and SD in either [-0.33, 0.33] during the time window indicating that the students were each participating in taking actions on the concept map, but on different concepts.

Learning performance: Correctness of the concept map

The learning performance for this activity is the correctness of the concept map. We asked two domain experts to create a map using the same 14 concepts. All the participant maps were compared against this expert map. We followed the following map-evaluation scheme: 1) 2 points for correct link and correct label; 2) 1 point for correct link and no label; and 3) 0.5 point for correct link and incorrect correct label. We added the points for each link between all the concepts and that was the dyad's performance score. Finally, we applied a median split to divide the dyads into high and low performance groups.

Data analysis

To examine the direct relationship between eye tracking measures and performance, dialogue codes, and DoL, we used a set of ANOVAs. We tested for the normality and homoscedasticity conditions using Shapiro-Wilk and Bausch-Pegan tests, respectively. In the case where the normality was violated, we normalized the data, and in cases where the homoscedasticity was violated, we used a Welch correction. We also tested for the pairwise interaction for all the variables using ANOVA. For the post hoc pairwise tests we applied Bonferroni corrections. We also computed the Cohen's d as the effect size for each ANOVA calculation. According to Cohen effect sizes can be low (below 0.2) medium (0.2 - 0.8) and high (above 0.8).

Results

JVA and JME relation with performance levels

To answer our first research question, we analysed the relationship between JVA and JME and the performance level of our dyads on their concept map. For both measures, we observed significant associations with performance level (see Figure 2, left column). The JVA for high performing dyads is significantly higher than the JVA for the low performing dyads ($F(1,38) = 18.67$, $p < .0001$, $d = 0.65$). Similarly, the JME for high performing dyads is significantly higher than the JME for the low performing dyads ($F(1,35.27) = 23.91$, $p < .0001$, $d = 0.81$).

JVA and JME relation with other process variables

To answer the first part of our second research question, we investigated how JVA and JME relate to the other collaborative process measures (i.e., dialogue content and division of labor). We observed significant associations between the DoL strategies with both JVA ($F(2,37) = 25.21$, $p < .0001$, $d = 0.87$) and JME ($F(2,32.45) = 8.29$, $p < 0.01$, $d = 0.24$). As seen in the middle column of Figure 2, JVA is highest when the students are engaged in role division compared to concurrent ($F(1,38) = 16.89$, $p < .01$, $d = 0.58$) or task ($F(1,38) = 27.49$, $p < .0001$, $d = 0.92$). Additionally, JVA is higher for a concurrent division than a task ($F(1,38) = 24.35$, $p < .001$, $d = 0.84$). Similarly, JME is highest for a role division compared with concurrent ($F(1,35.59) = 17.01$, $p < .01$, $d = 0.62$) or task ($F(1,31.35) = 28.33$, $p < .001$, d = 0.95), and task is also lower than concurrent ($F(1,33.24) = 22.43$, $p < .001$, $d = 0.78$).

Further, we found a significant relationship between both JVA and JME and the dialogue codes. In the right column of Figure 2, we see that JVA is significantly higher during concept-map dialogues than knowledge dialogues ($F(1,38) = 11.17$, $p < .001$, d = 0.38). We found the opposite for JME with JME being significantly higher during knowledge dialogues than concept-map dialogues ($F(1,36.83) = 31.29$, $p < .0001$, d = 1.03).

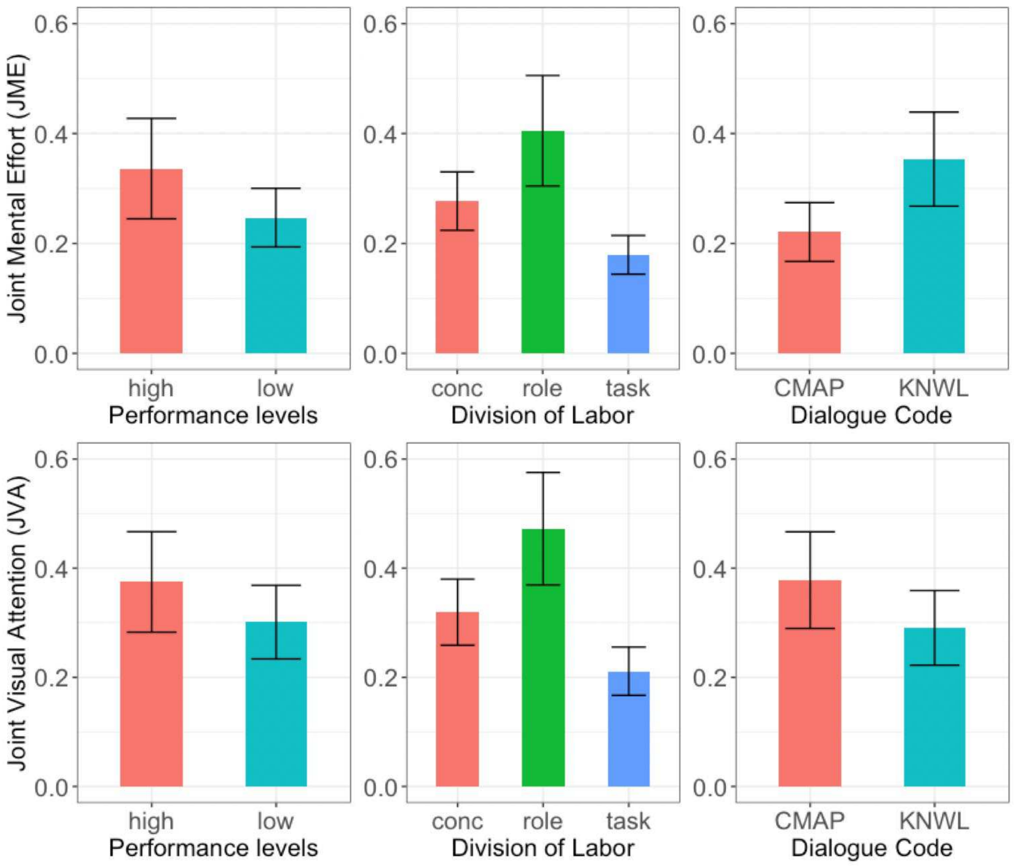


Figure 2. Comparing Joint Mental Effort (top panels) and Joint Visual Attention (bottom panels) across the different categories of Performance (left, high/low), division of labor (middle, concurrent/role/task), and dialogue codes (right, CMAP/KNWL).

Interaction effect between process variables and performance on eye-tracking

Finally, to answer the second part of our second research question, we analysed the impact of the interactions between the different process variables and their relation with the performance outcome. Concerning the interaction effects on JVA, we found a significant interaction between the performance levels and DoL strategies on JVA ($F(1,38) = 20.79$, $p < .0001$, $d = 0.72$). As we can see in Figure 3, the low performing students have a relatively stable JVA across the three DoL strategies. On the other hand, the high performing students fluctuated to have no significant difference with concurrent, higher JVA for role ($F(1,35.56) = 13.11$, $p < .0001$, $d = 0.47$), and lower JVA for task (d = $F(1,32.72) = 9.38$, $p < .0001$, $d = 0.38$) compared to the lower performing students. In terms of dialogue, we did not observe any interaction effect of performance and the dialogue category (CMAP/KNWL) on the JVA.

As with the JVA, we found an interaction between performance levels and DOL strategies for JME ($F(1,38) = 9.56$, $p < .001$, $d = 0.29$). There is not a significant difference between high and low performing dyads during concurrent and task divisions, but the difference is significant between high and low performing students during role divisions ($F(1,22.23) = 11.23$, $p < .0001$, $d = 0.39$) as seen in Figure 3. We did not observe a significant interaction between performance levels and the dialogue codes (CMAP/KNWL) on JME.

Discussion and conclusion

In this paper, we aimed to explore alternate measures of the collaboration processes, namely, ones gathered through dual eye tracking. As with other multi-modal studies, we are interested in the additive property that analysing the collaborative process from multiple perspectives can provide. With that in mind, in this discussion, we will present different interpretations of the data and how these interpretations are narrowed down as we add new measures, indicating the additive property of the measures rather than providing a set of proxies.

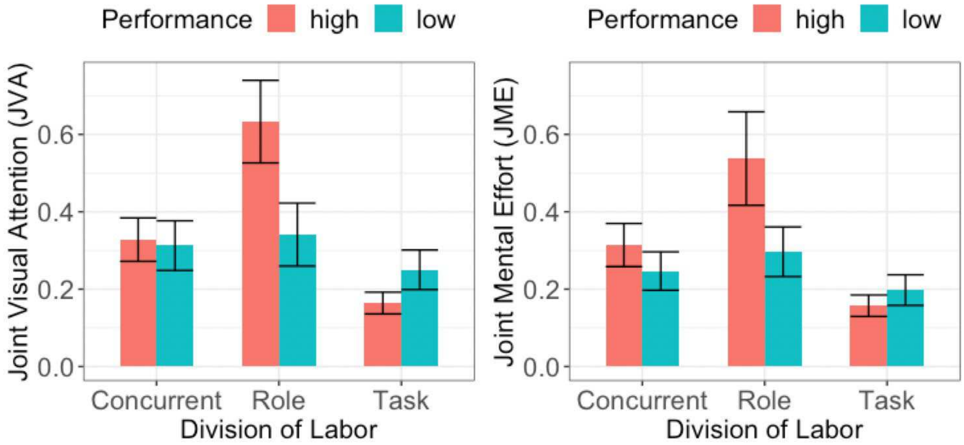


Figure 3. Interaction effect of performance levels and division of labor strategies on joint visual attention (left) and joint mental effort (right).

In terms of our first research question, how did the eye tracking measures relate to student performance, we confirmed our hypothesis that both higher JVA and JME were positively related with higher performance. This finding provides more of a confirmation of previous findings (Sangin et al., 2011; Jermann and Nuessli, 2012) than necessarily providing a new insight on its own. However, we would like to highlight that although researchers have studied the impact of individual cognitive load extensively (Amadieu et al, 2009, Kalyuga, 2011), using a joint measure to assess the collaborative process is relatively new. What this means is that students that are putting in the same amount of mental effort at the same time are more likely to perform well, and it is not just about the amount of mental effort of an individual student. The positive relation between JME and performance in a task where the transactive actions are taking place (division of labor, communication, and coordination of activities, Popov et al., 2017), gives an indication towards JME being a decent proxy for collaborative cognitive load. Nonetheless, it is an early indication and further studies are required for generalizability.

In terms of our second research question, how do JVA and JME relate to other process variables and what is the interaction when we include performance, it is interesting to discuss the results in terms of the division of labor measure. First, let's look at the concurrent division. Recall from our description of this measure that the concurrent division occurs when each member of the dyad is working on the same concepts in the same time window (although not necessarily trying to take the same actions). From our analysis, we found that during concurrent division, students had both JVA and JME measures somewhere between those of the role and task divisions. This might be expected as the students are looking in the same general area but, as they are doing separate actions, this overlap does not mean that they are necessarily working together. However, because the JME is also in between, it is unlikely that one student is just doing aesthetic changes while the other is enhancing the concept diagram. Although at the surface level, we may want to classify concurrent division as weak collaboration, as the students are doing separate actions, the eye tracking measures indicate that this is not always the case and warrants further investigation as to what occurred in the collaboration process prior that led to this division of labor before determining that an intervention is needed.

In contrast, during the role division, one student is doing all of the actions during a time frame. This pattern may be due to one partner free-riding (Le et al., 2018) or due to a productive driver/navigator collaboration (Bryant et al., 2006). If the students are focused on the same thing (high JVA) and are putting in the same mental effort (high JME), it is more likely to be a productive collaboration than a student free-riding. Although we found high JVA and JME in general during the role division, this was not the case when we took into account student performance. In this case, low performers had significantly lower JVA and JME, indicating that this may be instances of free-riding. Unlike with the concurrent collaboration, the use of roles is often considered as a productive collaboration script (King, 1999). Nonetheless, roles alone do not guarantee interdependence, and these may be moments for a clear intervention.

During the task division, students are working on different concepts during the same time window. As a first interpretation, this may mean that the students have divided the work evenly and are each working on a different part. However, it may also indicate that one student is doing the majority of the work while the other is making aesthetic changes – like when one participant writes a paper and the other corrects typos. From the JVA, we cannot differentiate these actions, as the JVA is low, as expected, due to the students working on different parts of the map. The JME can provide more insight though. We might expect the JME to be high if the students have an equal division of labor. In our case, we found the JME to be low, most likely indicating that there was not an equal divide. Perhaps more surprisingly, the JME was not different for high and low performers, but the JVA was. This may indicate that although in both cases the division of labor may not have been even, the high performers may have had more confidence carrying out their tasks independently. Future work would be needed to assess the exact task division and how these actions fit into the students' broader collaboration processes.

Finally, we observe that the JVA and JME have an opposite relationship with the dialogue, i.e., whether the dyad is talking about the interface or the domain knowledge. JVA is higher for the interface-based dialogues while the JME is higher for the knowledge dialogues. This indicates the complementary nature of two gaze measurements. The JVA is higher when there is strong visual support to ground the verbal references and JME is higher when the discussion is focused on domain knowledge. If we were to ignore one of these measurements, we would have received only half the picture (either in terms of attention management or in terms of effort management). Moreover, the high performing dyads have both JVA and JME that are higher than the low performing dyads, showing that, in terms of both their attention and effort management, the high performing dyads have more of an equal participation than the low performing dyads. This could possibly lead to better task-performance at the end of the collaborative session.

Although this paper presents a first analysis of how different types of eye tracking measures can shed light on the collaborative process, there are still several limitations. First, we only have an end performance measure. It is not clear if the collaborative process patterns we see are due to the knowledge of the students when they begin or if these actions led to better learning thereby leading to a better performance. Second, there are many combinations of process variables and temporal aspects that we did not explore in this paper that would provide further insights into the collaborative process and how it impacts JVA and JME. Due to space, we could not include them all.

In this paper, we aimed to deepen our understanding of what effective collaboration looks like through the assessment of dual eye tracking measures. We found that an effective collaboration is not necessarily a one-size-fits-all where a single metric can be used to judge the quality of the collaboration. Further, we found that joint mental effort can provide additional information than joint visual attention alone to better assess the collaborative process, contributing to the use of dual eye tracking methodology.

References

Amadieu, F., Van Gog, T., Paas, F., Tricot, A., & Mariné, C. (2009). Effects of prior knowledge and concept-map structure on disorientation, cognitive load, and learning. *Learning and Instruction, 19*(5), 376-386.

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 the *Procs. of the 36th Annual Conf. of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Brennan, S. E., Chen, X., Dickinson, C. A., Neider, M. B., & Zelinsky, G. J. (2008). Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition, 106*(3), 1465-1477.

Brown, A. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of Learning Sciences, 2*(2), 141-178.

Bryant, S., Romero, P., & du Boulay, B. (2006). The collaborative nature of pair programming. In *International Conference on Extreme Programming and Agile Processes in Software Engineering* (pp. 53-64). Berlin, Heidelberg: Springer.

Cherubini, M., Nüssli, M. A., & Dillenbourg, P. (2008). 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).

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

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).

D'Angelo, S., & Gergle, D. (2016). Gazed and confused: Understanding and designing shared gaze for remote collaboration. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 2492-2496).

Duchowski, A. T., Krejtz, K., Krejtz, I., Biele, C., Niedzielska, A., Kiefer, P., ... & Giannopoulos, I. (2018). The index of pupillary activity: Measuring cognitive load vis-à-vis task difficulty with pupil oscillation. In *Procs. of the 2018 CHI Conf. on Human Factors in Computing Systems* (pp. 1-13).

Hausmann, R. G., Chi, M. T., & Roy, M. (2004). Learning from collaborative problem solving: An analysis of three hypothesized mechanisms. Proceedings for *26nd Annual Conference of the Cognitive Science society* (pp. 547-552). Mahwah, NJ: Erlbaum.

Jermann, P. (2004). *Computer Support for Interaction Regulation in Collaborative Problem-Solving* [Unpublished doctoral dissertation]. l'Université de Genève, Switzerland.

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

Johnson, D. W., & Johnson, R. T. (1987). Research shows the benefits of adult cooperation. *Educational leadership, 45*(3), 27-30.

Joshi, M., & Rosé, C. P. (2007). Using transactivity in conversation for summarization of educational dialogue. In *Workshop on Speech and Language Technology in Education*.

Kalyuga, S. (2011). Cognitive load theory: How many types of load does it really need? *Educational Psychology Review, 23*(1), 1–23.

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.

Kirschner, F., Paas, F., & Kirschner, P. A. (2009). Individual and group-based learning from complex cognitive tasks: Effects on retention and transfer efficiency. *Computers in Human Behavior, 25*, 306–314.

Le, H., Janssen, J., & Wubbels, T. (2018). Collaborative learning practices: teacher and student perceived obstacles to effective student collaboration. *Cambridge Journal of Education, 48*(1), 103-122

Lemaignan, S., & Dillenbourg, P. (2015). Mutual modelling in robotics: Inspirations for the next steps. In *10th ACM/IEEE Intl. Conf. on Human-Robot Interaction (HRI)* (pp. 303-310). IEEE.

Lou, Y., Abrami, P. C., & d'Apollonia, S. (2001). Small group and individual learning with technology: A meta-analysis. *Review of educational research, 71*(3), 449-521.

Mercer, N. (2008). The seeds of time: Why classroom dialogue needs a temporal analysis. *The Journal of the Learning Sciences, 17*(1), 33–59.

Olsen, J., Sharma, K., Aleven, V., & Rummel, N. (2018). Combining gaze, dialogue, and action from a collaborative intelligent tutoring system to inform student learning processes. Proceedings of *the 13th International Conference of the Learning Sciences* (pp. 689–696). London, UK: ISLS.

Olsen, J. K., Sharma, K., Rummel, N., & Aleven, V. (2020). Temporal analysis of multimodal data to predict collaborative learning outcomes. *British Journal of Educational Technology, 51*(5), 1527-1547.

Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007). The art of conversation is coordination. *Psychological science, 18*(5), 407-413.

Rummel, N., Deiglmayr, A., Spada, H., Kahrimanis, G., & Avouris, N. (2011). Analyzing collaborative interactions across domains and settings: An adaptable rating scheme. In *Analyzing interactions in CSCL* (pp. 367-390). Springer, Boston, MA.

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.

Sangin, M. (2009). *Peer Knowledge Modeling in Computer Supported Collaborative Learning* [Unpublished doctoral dissertation]. École polytechnique fédérale de Lausanne, Switzerland.

Schneider, B. (2019). Unpacking collaborative learning processes during hands-on activities using mobile eye-trackers. Proceedings of *the 13th International Conference Computer-Supported Collaborative learning* (pp. 41-48). Lyon, France: ISLS.

Schneider, B., & Pea, R. (2015). Does seeing one another's gaze affect group dialogue? A computational approach. *Journal of Learning Analytics, 2*(2), 107-133.

Sharma, K., Jermann, P., Dillenbourg, P., Prieto, L. P., D'Angelo, S., Gergle, D., ... & Rummel, N. (2017). Proceedings of *the 12th International Conference Computer-Supported Collaborative learning*. Philadelphia, PA: ISLS.

Sharma, K., Chavez-Demoulin, V., & Dillenbourg, P. (2018). Nonstationary modelling of tail dependence of two subjects' concentration. *The Annals of Applied Statistics, 12*(2), 1293-1311.

Sharma, K., Leftheriotis, I., & Giannakos, M. (2020). Utilizing Interactive Surfaces to Enhance Learning, Collaboration and Engagement: Insights from Learners' Gaze and Speech. *Sensors, 20*(7), 1964.

Starr, E. L., Reilly, J. M., & Schneider, B. (2018). Toward Using Multi-Modal Learning Analytics to Support and Measure Collaboration in Co-Located Dyads. Proceedings of *the 13th International Conference of the Learning Sciences* (pp. 448–455). London, UK: ISLS.