Dyadic joint visual attention interaction in face-to-face collaborative problem-solving at K-12 Maths Education: A Multimodal Approach

Chiao-Wei Yang, Mutlu Cukurova, Kaska Porayska-Pomsta

UCL Knowledge Lab, London, UK

{chiao.yang.17}@ucl.ac.uk

Abstract. Collaborative problem-solving (CPS) is an essential skill in the workplace in the 21st century, but the assessment and support of the CPS process with scientifically objective evidence are challenging. This research aims to understand in-class CPS interaction by investigating the change of a dyad's cognitive engagement during a mathematics lesson. Here, we propose a multimodal evaluation of joint visual attention (JVA) based on eye gazes and eye blinks data as non-verbal indicators of dyadic cognitive engagement. Our results indicate that this multimodal approach can bring more insights into students' CPS process than unimodal evaluations of JVA in temporal analysis. This study contributes to the field by demonstrating the value of nonverbal multimodal JVA temporal analysis in CPS assessment and the utility of eye physiological data in improving the interpretation of dyadic cognitive engagement. Moreover, a method is proposed for capturing gaze convergence by considering eye fixations and the overlapping time between two eye gazes. We conclude the paper with our preliminary findings from a pilot study investigating the proposed approach in a real-world teaching context.

Keywords. Collaborative problem-solving. Multimodal learning analytics. Joint visual attention. Cognitive engagement. Temporal analysis. Eye-tracking.

1 Introduction

1.1 A need for scientifically objective evidence for CPS process assessment

CPS is an essential skill in the 21st-century workplace [1] and regardless of where CPS sits in the curriculum, teachers or educators are expected to equip students with this competence. This research attempts to create a multimodal temporal analysis of dyadic cognitive engagement as evidence for the analysis of students' cognitive engagement behaviours in CPS. Because interdependence is a key feature of CPS, a dyad is regarded as the unit of analysis. We argue that changes in levels of joint visual attention (JVA) may represent the embodiment of group cognition processes, and the temporal analysis of JVA makes the dyad's CPS process visible and comprehensible.

CEUR-WS.org/Vol-2902/paper6.pdf

Copyright © 2021 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

1.2 A short review of multimodal learning analytics (MMLA) in collocated collaboration

The depth and level of a team's engagement in face-to-face collaboration can only be understood through data and evidence, and multimodal learning analytics is a promising way of capturing and interpreting such data [2], [3]. Many recent studies have made use of indicators to measure the quality of collocated collaboration. These studies mainly focus on two types of indicators: social (verbal, non-verbal, and physiological) and epistemological (logs and ideas) [4]. However, more subtle indicators of internal cognitive states are rarely discussed in terms of assessing a team's collaboration performance. Furthermore, the theme of cognitive engagement is less common in MMLA research. As highlighted in a recent review of the field [5], there is a lack of MMLA studies investigating the association between the mode of gaze employed and the research theme involving engagement. There is also a lack of studies showing the association between gaze modality and teamwork in formal learning. This paper aims to explore what insights eye physiological data can provide in CPS assessment. It focuses on cognitive engagement as an indicator of a dyad's CPS performance.

The next section will explain engagement in terms of an engagement framework. It will highlight ways to measure dyadic engagement, including justification of using eye gaze and eye blinking data as indicators of dyadic cognitive engagement.

1.3 Halverson and Graham's (HG) Cognitive Engagement Framework

This paper follows Halverson and Graham's [10] definition of engagement, whereby cognitive engagement includes behavioural engagement. The researchers emphasised identifying engagement through cognitive and emotional indicators, arguing that external behaviours are "the outward displays of the mental and emotional energies that fuel learning" ([10], p.153). Even though cognitive engagement in this study comprises both behaviour and cognition, emotional engagement' is considered as beyond the scope of this study.

According to the HG framework, several factors indicate the quantity of cognitive engagement (attention, effort, persistence, and time spent on task) and a number of factors indicate the quality of cognitive engagement (cognitive strategy use, absorption/deep concentration, and curiosity). Since JVA measures attention, one of the factors concerning the quantity of cognitive engagement, it will be used as the proxy measure for the quantity of engagement in this research. To detect JVA data, eye gazes were measured according to Just & Carpenter's eye-mind hypothesis [11]. This is based on their observation that eye movements are closely linked to mental activity. In terms of the eye blinking rate (EBR: the number of eye blinks per minute), changes in EBR are used to interpret deep concentration, as an index for the quality of dyadic cognitive engagement. EBR has been studied in neuroscience and psychological research (e.g biological psychology). Even though several research studies indicate that an increased EBR correlates with higher dopamine (DA) levels [12], it is argued that blinking rates were determined by the 'task' rather than the dopaminergic state

[13]. Besides, there have been many studies that related eye blink rate with cognition, particularly in task difficulties [14-15], the attention required in tasks [16] or task engagement [16]. Evidence from studies mentioned above demonstrates that spontaneous blinking is suppressed to minimise the loss of visual information when the visual information is more important to a person. Although contexts differ in the papers discussed above, the evidence presented provides sufficient ground to establish the relevance of EBR as an indicator of absorption (deep concentration). This may potentially be explained as a person's need for high cognitive attention in order to experience focused concentration – a state of flow. Notably, absorption here does not refer to the act of paying attention. It means a "state in which people are so involved in an activity that nothing else seems to matter" (Csikszentmihalyi, 1990, p.4, cited in [10], p.156).

2 Research Problems & Research Questions

There are recent studies investigating the association between joint visual attention (JVA) and high-quality collaborative interactions of students [6-8,19]. However, based on the unimodal data, counting the number of joint eye gazes alone seems insufficient as a measure of the quality of collaboration [8]. The insufficiency of unimodal JVA data to fully represent collaboration is also echoed by Siposova and Carpenter [9]. The authors argued that the nature of social attention is complex since the jointness of attention comes in degrees rather than as arbitrary, discrete and uniform events. Inspired by gaps in MMLA and the theoretical propositions on the JVA's temporal nature, we propose the research questions below.

RQ1. To what extent can eye blinking physiological data increase our understanding of dyadic cognitive engagement in the CPS context?

RQ2. What insights can multimodal JVA data generate when adopted in the measurement of dyadic CPS competence in face-to-face, K-12, Maths learning contexts?

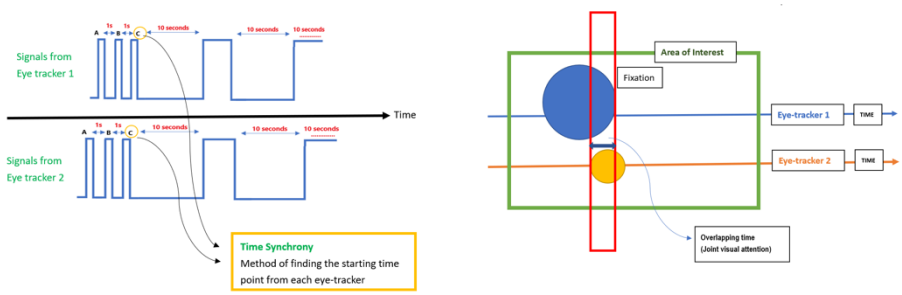
In this research, "unimodal-based JVA" refers to levels of joint visual attention identified via counting the frequency of joint eye gazes in a dyad. Multimodal-based JVA refers to levels of JVA identified by combining joint eye gaze data as well as an individual student's eye blinking rate (EBR) in a dyad. More details about techniques to capture each indicator will be discussed in the next section. All signals were collected from eye image videos of two mobile eye-trackers (Tobii Pro Glasses 2) and were synchronized using Tobii Pro Lab software.

3 Methodology

3.1 Multimodal data collection

A new approach to capturing joint visual attention

Gaze convergence is often used to measure a dyad's collaborative outcome or performance. In recent studies, there are two alternative measures of gaze convergence [6,19] which are commonly assumed when two subjects are looking at the same place at the same time. One is to capture a dyad's joint visual attention [19] and the other is to gauge the gaze similarity [6]. Both approaches have limitations. For instance, the use of fiducial markers for participants to glance at every time before collaboratively solving task problems possibly distracts participants from engaging in CPS activities [6,19]. Also, Schneider's approach [19] included short fixations within the arbitrarily defined distance (e.g., radius size 100 pixels) between two gaze points to be considered as a moment of joint visual attention. That would reduce the JVA's detection accuracy since the time is needed for the brain and eye to process what is seen [20]. Regarding the gaze similarity measurement used by Sharma and other researchers [6], despite considering an individual's eye fixation, they don't consider the overlapping time when two eye gazes meet together in the same area (Area of Interest, AOI). This shows the limitations of capturing accurate JVA because of the essential "jointness" idea in its measurement. The following paragraphs present a proposed approach to JVA measurement (See Fig.1 A, B).



A. Approach for time synchrony B. Gaze convergence detection

Fig. 1. Graph A: The use of eye-tracking software to detect identical starting points from two eye- tracking devices for time synchrony. Graph B: One moment in time captured to represent eye gazes (Fixation) from both participants remain on one area (Area of Interest) on a shared screen. The overlapping area is defined as joint visual attention in a dyad.

Eye blinking data detection

The initial images of the eyes provided by the built-in camera in an eye tracker are translated into images of pupil location on the software. The image is inverted so that white indicates the pupil location during one occlusion, meaning the time between blinks. If the computer registers a white mark, then this means the eye pupil is seen. The eye must be open and the time between blinks can be measured. A lack of white mark, suggests that the eye is closed. The time between the white mark appearing and disappearing is treated as one blink. The algorithm accurately detected above 88.0% of all blinks identified by manual coding of eye image videos in the pilot study.

3.2 Temporal analysis.

Temporal analysis is proposed in this research as the most relevant approach for examining CPS competence for three reasons. First, the situational context of CPS and the in-process measurement play a more significant role (as in formative assessment, rather than summative assessment) in CPS assessment. CPS competence is a dynamic process heavily dependent on context and temporal dimensions. Second, the temporal analysis allows the presentation of visual snapshots of a dyad's CPS learning process via nonverbal eye interaction in collocated classroom collaboration. In particular, it can support locating some key moments of learning that have been missed in both quantitative and qualitative methodology. For instance, the use of frequency and average measurement, as well as self-report interviews, are challenging to be used to reflect changes in dyadic engagement over time. Third, the temporal analysis aligns with the characteristics of cognitive engagement and its indicators. For instance, the way engagement can vary in intensity and duration [21], meaning that engagement can be measured on a continuum (a single dimension of engagement ranging from high to low) rather than a binary categorization (engaged or disengaged).

3.3 Pilot Study

A secondary school math teacher recruited four 13-year-old students to be paired into two groups. One group had high average math grades (average math grades ranked first and third, dyad B). The other one had low average math grades (average math grades were second-to-last, dyad A). The two pairs then took turns participating in the experiment. During the 20-minute classroom observation, a dyad of two shared a tablet and did math exercises on the learning platform. The students collaboratively solved one-variable linear equations. Once they agreed on the answers, they submitted them to the system by clicking the submit button, and also wrote them down on the shared worksheet. Each student wore a mobile eye-tracker (Tobii G2) during the CPS activity. In the first stage of the study (approximately the first 5 mins), the students were presented with four math questions. In the second stage (around 12 mins), the students were given another 8 math questions. A dyad's engagement was measured in terms of the joint cognitive engagement state through investigating a dyad's joint visual attention and eye blinking rates.

4 Results

There are two preliminary findings from these data. First, a dyad's concentration state in a certain moment may be predicted by observing synchronised EBR patterns in the CPS process (Figure 2). Secondly, the multimodal JVA data with a temporal analysis seemed promising, as it provided information about not only the level of dyadic engagement but also the frequency of a dyad's highest /lowest engaged states (e.g., peaks and troughs) during the CPS process (Figure 3, C, D).

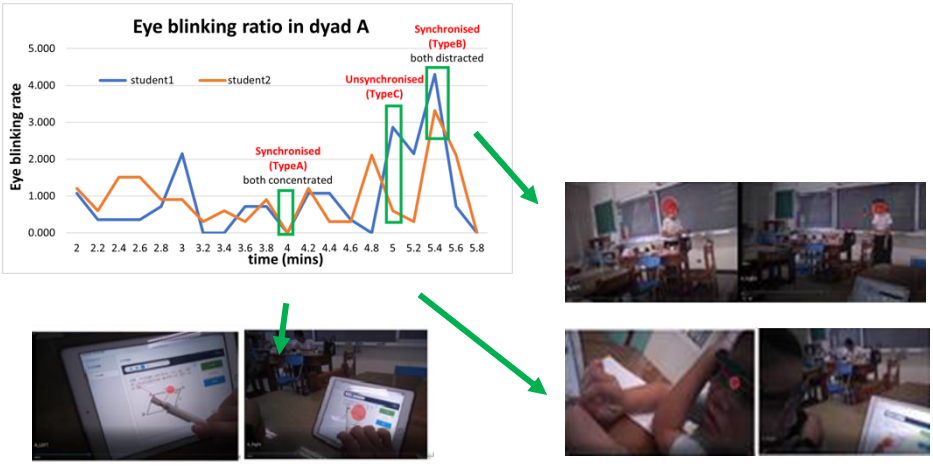


Fig. 2. The change of two students' eye-blinking rate, in a dyad over time in the pilot study. Green columns indicate three different types of EBR patterns in 4 mins, 5.1 mins, and 5.5 mins timepoints respectively. Video snapshots were used to demonstrate three types of EBR patterns.

4.1 Synchronised eye blinking data in a dyad's CPS process

Figure 2 displays the pattern of the change in eye blinking ratio over time of two students. A pair (dyad A) synchronized (type A & B) and unsynchronized patterns (type C) are illustrated via green boxes. Two students collaboratively finished the task for around 6 minutes (Figure 4), resulting in a synchronized EBR type A pattern being observed at the 4th min (Figure 2). Students didn't appear to concentrate on the assigned task when unsynchronized type C and synchronized type B patterns were observed (Figure 2), as illustrated at the 5.1 and 5.5 mins timepoints, respectively.

4.2 Unimodal and multimodal JVA, temporal analysis graphs

In figures 3 and 4, the pairs of students' JVA in unimodal (as measured only with eye gaze data) and multimodal (as measured with a combination of eye gaze and blink data) are graphically presented. Despite both unimodal and multimodal JVA graphs showing joint visual attention between a pair of students over time, the indicators used for these measurements were different. The unimodal JVA graph (Fig. 3 A, B) measured levels of JVA, without EBR included, whereas the multimodal JVA graph (Fig. 3 C, D) included combined EBR from two students. The results of unimodal and multimodal graphs differed significantly (Fig. 3). For instance, between timepoint 3.6 and time point 3.8 mins in the activity, levels of JVA were expected to be lower because a dyad A was not focusing on solving math questions. However, levels of JVA were still relatively high in the unimodal data graph (Fig. 3A). In addition, in dyad B, at the moment between timepoint 5.6 mins and timepoint 6.2 mins, the expected trend was from a trough to a peak due to the fact that the dyad was submitting an answer and then actively discussing the next math question right away. However, the levels

of JVA in the unimodal data graph (Fig 3B) were very high during the same period. In Figure 4, video snapshots showed that peaks and troughs in the multimodal JVA graph were more accurately representing the CPS interactions (high/ low engagement) in dyad A.

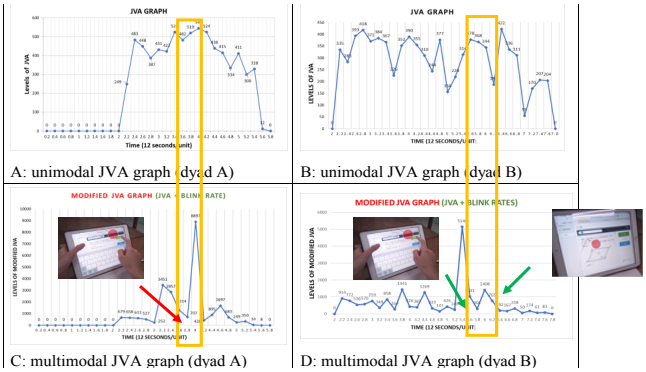


Fig. 3. (A, B) Unimodal and (C, D) multimodal evaluation over time (secs). Yellow boxes indicate in (A, C) dyad A, and (B, D) dyad B. Arrows indicate a dyad's behaviour in a CPS context. Dyad A has low academic performance, while dyad B has a high performance.

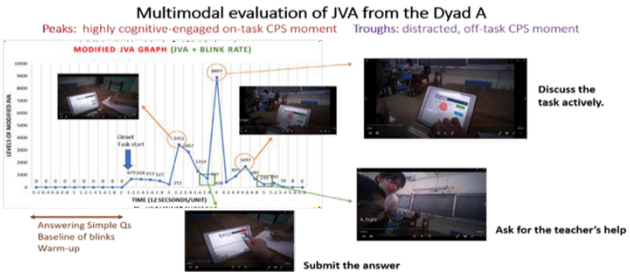


Fig. 4. An example of a multimodal JVA graph illustrating how moments in dyad A's CPS interactions are correctly reflected in the graph

5 Discussion

The present study aimed to identify how to plot JVA over time to accurately represent a pair of students' behaviours to solve math problems collaboratively. Additionally, to reflect the diminishing levels of dyadic cognitive engagement when only one student uses the shared tablet to enter and submit answers. This translates to high levels of JVA when both students in a pair make mental efforts during the CPS process, and low levels of JVA when only one student enters information and sends it to the system. To address this, video snapshots were used to determine which JVA graphs more accurately represented a dyad's different moments of engagement.

In addressing RQ1, our pilot study results demonstrated that eye blinking data can be useful to increase researchers' (or potentially other stakeholders such as learners

and teachers) understanding of dyadic cognitive engagement in the CPS context. The change in the EBR indicated the dyad's concentration level, which may increase the accuracy of the JVA interpretations (levels of JVA in graphs) in dyadic interactions. One assumption of this methodology is that observing the number of eye blinks over time may allow the researcher to determine how students' states of absorption (deep concentration) change. In terms of interpreting eye blinking data, the lower the number of blinks, the more concentrated the learner was considered in a CPS context; conversely, the higher the number of blinks, the lower their concentration. When EBR data streams were added to unimodal JVA data streams based on eye gaze data, each dyad's key moments of intense concentration and frequency of peaks and troughs in the CPS process emerged in multimodal JVA graphs. Based on the observation of student behaviours from the video recordings of their behaviours and using it as the ground truth of their engagement, we concluded that the information generated from the EBR makes the non-verbal multimodal JVA temporal analysis graph more informative and accurate.

Regarding RQ2, exploring the insights that multimodal JVA data from eye gaze and eye-blinks can help us generate in the measurement of dyadic CPS competence in face-to-face, K-12, Maths learning contexts. We observed that eye blink data can bring in valuable information about students' deep concentration during their CPS process. Such insights can help us design AIED and Learning Analytics tools to improve children's CPS competence at K-12 schools. For instance, teachers can use the graphs presented in dashboards to identify peaks in the intensity of JVA and identify topics, exercises, or tasks that can incite more discussion in a dyad. They also can use the graphs to identify the frequency of highest and lowest engagement states to give appropriate interventions or to support students reflecting their CPS behaviours. It is important to note that the value of such graphs for researchers are highlighted here, but their potential for teachers would require significant design work involving teachers. So, our discussions about their value to teachers here are mainly to generate hypothesis to be studied in the future.

There are many possible explanations as to why multimodal data graphs can more accurately represent changes in dyadic cognition engagement compared with unimodal data graphs. Firstly, joint attention is dynamic, not arbitrary [9], therefore, only relying on unimodal joint eye gaze data cannot accurately reflect the quality of a pair's collaboration. Secondly, eye blinking frequency and joint eye gaze can be accurately captured using a mobile eye tracking device alone (See section 3.1). Finally, the definition of cognitive engagement (higher learning construct) was defined precisely before selecting indicators (lower data streams), and indicators were based on the aforementioned literature findings. These factors are crucial to enhance the accuracy of data collection, as arbitrarily selected indicators and vague definitions of complex engagement, may lead to inaccurate data interpretation.

Although multimodal data could bring insights to the learning analytics field [22, 23], it is not the purpose of this research to argue that multimodal data is better than unimodal data. Rather, we emphasise the eye blink data as an additional modality to eye gaze data can contribute new information to our evaluations of CPS, with the

fused data offering different perspectives about the students' JVA during CPS activities.

In order to meaningfully interpret the value of our proposed research, a significant number of comparative participant pairs need to be recruited, and entire session data should be analysed for the ecological validity of our interpretations. Moreover, study compliance was impacted by the intrusive nature of the eye-tracking devices during CPS activities, resulting in lost data. This is a significant issue that needs to be addressed in the planning of future investigations.

6 Conclusion and Future research

This research uses eye physiological data, eye blinks, and eye gaze behaviours to provide a multimodal interpretation of the dyad's JVA during CPS activities. We drew from the field of cognitive neuroscience to form the initial hypothesis that eye blink data can be a valuable data input in multimodal learning analytics approaches to interpreting JVA. The research proposed and piloted here has the potential to contribute to the literature with a new technique to capture joint visual attention (JVA) in collocated collaboration, and to demonstrate that a dyad's cognitive engagement change can be accurately observed by measuring levels of JVA on the temporal analysis with a multimodal approach.

In future work, the potential of the 'synchronized eye gazes and eye blinking rate' multimodal data as a parameter to measure cognitive engagement with AIED systems used in in CPS contexts should be further investigated. Within a CPS context, the AI system does not limit itself to face-to-face collaborative interactions between students. These situations can be any of the following: 1) the collaborative relationship between a virtual agent and a student in an intelligent tutoring system, 2) the CPS interaction between a learner and a robot in a human-robot interaction, or 3) the CPS interaction between two students in a virtual environment. Due to different dynamics of each particular context, the value of JVA and EBR to be applied in AIED systems should be studied separately in future research.

References

1. Graesser, A.C., Fiore, S.M., Greiff, S., Andrews-Todd, J., Foltz, P.W., Hesse, F.W.: Advancing the Science of Collaborative Problem Solving. Psychological Science in the Public Interest 19, 59-92 (2018)
2. Blikstein, P., Worsley, M.: Multimodal Learning Analytics and Education Data Mining: using computational technologies to measure complex learning tasks. Journal of Learning Analytics 3, 220-238 (2016)
3. Cukurova, M., Giannakos, M., Martinez-Maldonado, R.: The promise and challenges of multimodal learning analytics. British Journal of Educational Technology 51, 1441-1449 (2020)
4. Praharaj, S., Scheffel, M., Drachsler, H., Specht, M.: Multimodal Analytics for Real-Time Feedback in Co-located Collaboration. pp. 187-201. Springer International Publishing, (2018)
5. Sharma, K., Giannakos, M.: Multimodal data capabilities for learning: What can multimodal data tell us about learning? British Journal of Educational Technology 51, 1450- 1484 (2020)

6. Sharma, K., Leftheriotis, I., Giannakos, M.: Utilizing Interactive Surfaces to Enhance Learning, Collaboration and Engagement: Insights from Learners' Gaze and Speech. Sensors 20, 1964 (2020)
7. Jermann, P., Mullins, D., Nüssli, M., Dillenbourg, P.: Collaborative Gaze Footprints: Correlates of Interaction Quality. . In: Connecting Computer-Supported Collaborative Learning to Policy and Practice, pp. 184-191. International Society of the Learning Sciences. (2011)
8. Bryant, T., Radu, I., Schneider, B.: A qualitative analysis of joint visual attention and collaboration with high-and low-achieving groups in computer-mediated learning. In: Proceedings of the 13th International Conference on CSCL pp. 923-924. International Society of the Learning Sciences, (2019)
9. Siposova, B., Carpenter, M.: A new look at joint attention and common knowledge. Cognition 189, 260-274 (2019)
10. Halverson, L.R., Graham, C.R.: Learner engagement in blended learning environments: A conceptual framework. Online learning 23, 145-178 (2019)
11. Just, M.A., Carpenter, P.A.: A theory of reading: From eye fixations to comprehension. Psychological Review 87, 329-354 (1980)
12. Jongkees, B.J., Colzato, L.S.: Spontaneous eye blink rate as predictor of dopamine-related cognitive function—A review. Neuroscience & Biobehavioral Reviews 71, 58-82 (2016)
13. Van der Post, J., de Waal, P.P., de Kam, M.L., Cohen, A.F., van Gerven, J.M.A.: No evidence of the usefulness of eye blinking as a marker for central dopaminergic activity. Journal of Psychopharmacology 18, 109-114 (2004)
14. Drew, G.C.: Variations in Reflex Blink-Rate during Visual-Motor Tasks. Quarterly Journal of Experimental Psychology 3, 73-88 (1951)
15. Oh, J., Jeong, S.-Y., Jeong, J.: The timing and temporal patterns of eye blinking are dynamically modulated by attention. Human Movement Science 31, 1353-1365 (2012)
16. Stern, J.A., Walrath, L.C., Goldstein, R.: The Endogenous Eyeblink. Psychophysiology 21, 22-33 (1984)
17. Fairclough, S.H., Venables, L.: Prediction of subjective states from psychophysiology: A multivariate approach. Biological Psychology 71, 100-110 (2006)
18. Skinner, E., Furrer, C., Marchand, G., Kindermann, T.: Engagement and disaffection in the classroom: Part of a larger motivational dynamic? Journal of Educational Psychology 100, 765- 781 (2008)
19. Schneider, B.: Unpacking Collaborative Learning Processes During Hands-on Activities Using Mobile Eye-Trackers. In: 13th International Conference on CSCL, pp. 41-48. International Society of the Learning Sciences, (2019)
20. Munn, S.M., Stefano, L., Pelz, J.B.: Fixation-identification in dynamic scenes: comparing an automated algorithm to manual coding. Proceedings of the 5th symposium on Applied perception in graphics and visualization, pp. 33–42. Association for Computing Machinery, Los Angeles, California (2008)
21. Fredricks, J.A., Blumenfeld, P.C., Paris, A.H.: School Engagement: Potential of the Concept, State of the Evidence. Review of Educational Research 74, 59-109 (2004)
22. Cukurova, M., Kent, C., & Luckin, R. Artificial intelligence and multimodal data in the service of human decision-making: A case study in debate tutoring. *British Journal of Educational Technology*, *50*(6), 3032-3046. (2019)
23. Giannakos, M. N., Sharma, K., Pappas, I. O., Kostakos, V., & Velloso, E. (2019). Multimodal data as a means to understand the learning experience. International Journal of Information Management, 48, 108-119.