Zoom lens: An MMLA framework for evaluating collaborative learning at both individual and group levels

Qi Zhou [1], and Mutlu Cukurova [1]

[1] University College London, 23-29 Emerald St, London, WC1N 3QS, UK

Abstract

Collaborative learning (CL) has been considered to have great potential to help learners with their cognitive development, problem-solving skills, and social interaction skills. However, the analysis and support of this complex process, which requires meaningful interactions among learners at cognitive, social, emotional and regulatory levels, need holistic investigations at multiple dimensions. In recent years, multimodal learning analytics (MMLA) used to provide such multidimensional investigations to generate insights into the process of CL as well as the automatic analysis of certain aspects of CL. Previous research studies suggested that the analysis of collaboration should take both individual and group levels into consideration, yet most existing research only focuses on one level ignoring the other. Group-level information is needed for the success of a group to observe interindividual aspects of collaboration. In addition, learners may also need information from the individual level to help them be aware of their own performance and realise personal development in CL. Therefore, this paper suggests an MMLA framework developed from the concept of "from clicks to constructs" for supporting CL from both individual and group levels which aims to act as a "zoom lens" to provide information about the process of CL from multiple levels and dimensions.

Keywords

Face-to-face collaborative learning, Multimodal Learning Analytics

1. Introduction and Background

Collaboration has become an even more important skill in the twenty-first century since the need for thinking and working style in society has increasingly moved the emphasis from individual work to group work [16]. Collaboration requires the group members to share the authority for the actions as well as build consensus through working together [10]. Applying this philosophy generally as a way of living and interacting with other people, collaborative learning (CL) is a teaching and learning approach which encourages a group of learners to work together [18]. CL benefits learners from many different aspects, including learning how to resolve social problems [9], developing social interaction skills [3], building more positive heterogeneous relationships [26], encouraging diversity understanding [24], and helping learners to resolve differences in a friendly manner [10]. However, success in CL cannot be guaranteed by simply asking learners to work together [22]. CL is a complex process in which learners achieve development through meaningful interactions among individuals in the group.

In the past decade, learning analytics (LA) and multimodal learning analytics (MMLA) were used by researchers to explore CL. As a research field focusing on the measurement, collection, analysis and reporting of data about learners and the learning environment [19], LA provides opportunities for predicting collaboration performance [21], investigating the different mode of collaboration [13], and providing adaptive support for groups [11]. Extending the data sources of LA, MMLA collects, integrates, analyses, and interprets data from multiple modalities to gain a holistic understanding of the process of learning [5]. In the context of CL, it contributes to understanding the comprehensive

MILeS 23:The Third International Workshop on Multimodal Immersive Learning Systems, September 4, 2023, Aveiro, Portugal
EMAIL: qtnvqz3@ucl.ac.uk (Q. Zhou); m.cukurova@ucl.ac.uk (M. Cukurova);
ORCID: 0000-0002-4694-4598 (Q. Zhou); 0000-0001-5843-4854 (M. Cukurova);

behaviours and complex interactions exhibited by learners through different channels of data. For example, Ma et al., [12] collected video and audio data from pair coding activities and tried to use them to automatically detect impasse moments during collaborative problem-solving. To be more specific, linguistic features and acoustic features were extracted from the audio data while facial features were extracted from the video data. The impasse moments were manually annotated by two annotators through the textual transcripts of learners' dialogue during the collaboration. Then machine learning algorithms were used to explore the predictive models for detecting the impasse moment with different modalities of data. The result showed that the use of multimodal data might significantly increase the performance of the predictive model. It illustrated that multimodal data could contribute to providing information about learners' complex interactions during collaborative activities.

Most studies do indeed acknowledge two levels of subject considerations in CL, the individual level and the group level [1]. Which level should be assessed, that is, test scores for individual level versus quality of the projects for group level, is debated since the CL approaches developed in the 1970s [2]. To explore this issue, Webb [27] conducted research with 53 seventh-grade learners in a mathematics class to compare performance in small-group and individual assessment contexts. Learners were asked to solve mathematics problems both individually and collaboratively. The processes of collaboration were also observed and coded by researchers. The result showed that learners' individual competence might be different from the quality of their collaboration products. In contrast, the collaborative behaviours exhibited by learners during the processes of collaboration can provide more important and accurate information about learners' competence. It means that both individual-level and group-level evaluations need to be considered when analysing CL. For group-level evaluation, the processes of collaboration are more educational and meaningful than the quality of the collaboration product. Similarly, Janssen et al., [8] argued that conducting multilevel analysis is helpful in the analysis of CL. Given the fact that there are usually at least two levels of data (individual-level and group-level) when analysing CL, the datasets might be hierarchically nested, the variables might be non-independent, and there might be many different units of analysis. The application of multilevel analysis will help to provide a more comprehensive interpretation of CL.

In the field of Computer-Supported Collaborative Learning (CSCL), there is limited work focusing on both individual and group levels at the same time. On the one hand, many studies used group-level measurements, such as quality of the products, and human-observed collaboration quality, as their target constructs in the analysis. On the other hand, some researchers also stressed the importance of evaluating CL from individual level. For instance, Zhou et al., [31] suggested that providing individual behaviour might help learners to be aware of their own and others' contributions so that they can take action to improve their collaboration. Fernandez-Nieto and colleagues [6] used a story-telling approach to visualise learners' multimodal data and their group status to help them reflect on their performance in a nursing simulation activity. Yet, the visualisation is more like a report of "what has happened" rather than an analysis of "what has happened meaningfully in a collaboration context". This presented a challenge on how to incorporate the analysis of individual development in CL.

Therefore, this paper will present an MMLA framework which tries to take both group-level and individual-level into consideration when analysing CL. The rest of the paper will describe the details of the framework and introduce how it is implemented in a real-world face-to-face CL context.

2. The MMLA Framework

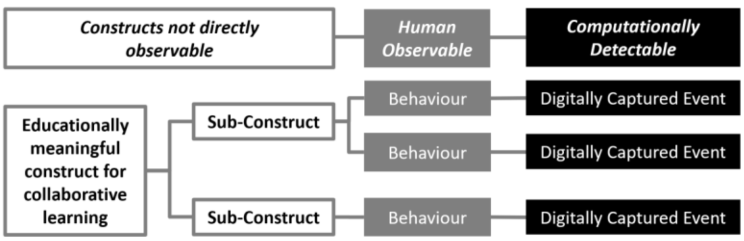

Fig. 1. From clicks to constructs [28]

The proposed framework is developed from the concept of "from clicks to constructs" which is coined to address the challenges of the differences between computational logged events and humanconceptualized activities in LA. In CSCL, Wise et al., [28] and Martinez-Maldonado et al.,[14] adapted and visualised in the context of CSCL (Fig. 1). Three key concepts were introduced in the graph, namely computationally detectable events, human observable behaviours, and constructs not directly observable. For instance, computationally detectable events are the data and features that can be automatically collected or generated by computer algorithms. Human observable behaviours refer to the behaviours exhibited by learners and can be observed by human beings with expertise in specific learning domains. The constructs that are not directly observable present the indicators that can be used to evaluate the quality of collaboration but cannot be directly reached through human observations. These constructs are usually based on the theory of collaborative learning. However, in this graph, the human observable behaviours were not described detailed enough and group level features are absent. There might be different types of generic individual behaviours (e.g. speaking, listening, and watching), specifically collaborative individual behaviours (e.g. Engaging in discussion, declaring knowledge, suggesting relevant ideas etc.), and collaboratively meaningful group interactions (e.g. Negotiating, referring and following other group members etc.). Each type of behaviour provides different information about the CL process. Therefore, there is a need to clarify what types of behaviours can be used to analyze CL and how they relate to each other.

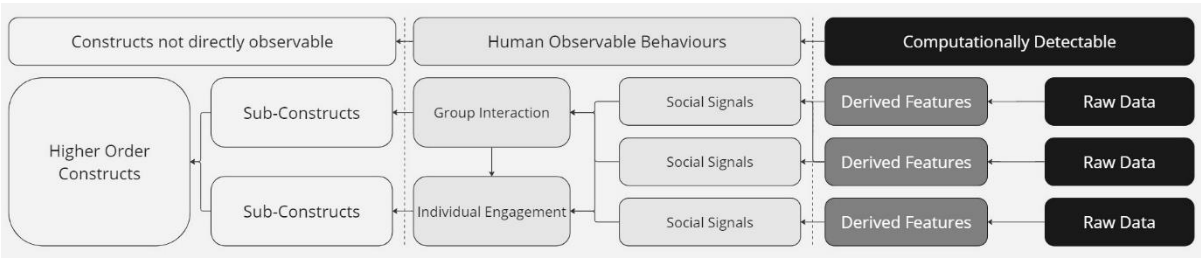

Fig. 2. Zoom Lens: An MMLA framework for multilevel analysis in CL

As Fig. 2. shows, The proposed framework adapted three key concepts from previous' work and a more detailed structure was presented.

2.1. Computationally Detectable Features

This layer includes the raw data and the derived features. Raw data is the multimodal data collected through different sensors such as video data, audio data, log data, and physiological data. The derived features are the measurements extracted from these multimodal data to be used for further analysis, such as linguistic features extracted from dialogue data, acoustic features extracted from audio data, gesture features and facial features extracted from video data, as well as the arousal features extracted from physiological data. Then these features will be used to detect the human observable behaviours. It is worth noting that the types of features were not "one-to-one" mapping with the behaviours at the next layer. Multiple types of features can be used to detect single behaviour, and one type of feature can also be used to detect multiple behaviours.

2.2. Human Observable Behaviours

Three types of behaviour markers have been defined in the layer of human observable behaviours. The first type is social signals. It refers to the behavioural cues (e.g. physical appearance, gestures and postures, face and eye behaviours, vocal behaviours, and environment) frequently used to understand and interpret human social states, such as emotion, attitude, physical interactions, and emblems [25]. Since these signals are the behaviours that learners directly exhibit but not interpret by the learning context, it can be easier to be detected the derived features with high accuracy through machine learning algorithms. For instance, gaze behaviours, speaking behaviours, and gestures can be detected from video and audio data. Physiological data can provide information about students' affections.

The second type of behaviour marker is group interaction. It refers to the interactions happening between the group members. It is used to describe the status of the whole group during the collaborative activity, such as interacting with peers through communication, resource management, and regulative behaviours [17]. This type of behaviour can be inferred from the gathering of each group member's social signals. For example, Zhou et al., [29] used an optimal matching approach to determine two different types of gaze interactions through individual group members' gaze behaviours.

With each individual's social signals and group interactions, the third type of behaviour, individual engagement, can be estimated. The individual engagement behaviours present whether learners participate in group interactions. For example, Cukurova et al., [4] developed a framework using nonverbal indexes of learners' physical interactivity to interpret the key parameters of learners' collaborative problem-solving competence. In the framework, each learner was coded with three different statuses (active, semi-active, and passive) based on their own basic behaviours (hands on objects, looking at active peers, etc.) and the group interaction contexts.

The relationship between these three types of behaviour markers can be built in different ways. On the one hand, a rule-based approach can be used to infer group interactions and individual engagement from social signals. It will rely on the CL researchers' interpretation of the relationship between these behaviours. This approach may contribute to building a transparent and explainable model for CL. On the other hand, algorithms from computer science can also be used to build the model to present the relationship between different types of behaviour markers. Generally, it can provide models with higher accuracy than the models build by rule-based approaches. But given the fact that most of these kinds of models are "black boxes", the interpretability and generalizability of the model are usually questioned in previous research [32]. The trade-offs between these two approaches need to be considered based on the aim of using MMLA in CL.

2.3. Constructs that are not directly observable

Constructs not directly observable are the indicators or measurements for evaluating the learning outcomes or individual development but cannot be directly captured by humans. The higher-order constructs can be assessed from sub-constructs based on educational theories [14]. For example, in a CL context, the higher-order construct can be learners' collaborative problems solving skills, and it can be assessed by the cognitive presence and social presence [7]. These sub-constructs are inferred from the behavioural markers from the last layer by traditional statistical methods [30], rule-based methods [4], process mining methods [29], and machine learning methods [20].

3. Implementation

This section will present a case study about implementing the proposed framework in a real-world CL context. The case study mainly focused on analysing collaborative learning from non-verbal behaviours. Only video and audio data were collected and analysed in this study. In the future, more modalities of data, and verbal behaviours analysis might be involved to provide a more comprehensive picture of collaborative learning.

3.1. Context

The study was conducted in a 10-week postgraduate module in the UK. Thirty-four learners were divided into groups of four or five based on their gender, learning and working experiences, cultural backgrounds, and expertise domains. During the course, learners were asked to work in groups to design a technological solution to address an educational challenge chosen by themselves. For each week, learners had to participate in a two-hour face-to-face session, including a one-hour tutorial and one-hour group discussion to work on their weekly collaborative group tasks. Before the face-to-face session, learners also need to read the literature related to the weekly topics, view the pre-recorded lectures, and join in the online asynchronous class debate via Moodle forum. During the whole week, learners had to reflect on their learning experiences individually and write them down via Google Docs. This study mainly focused on the data collected from the face-to-face discussion sessions.

3.2. Data collection

During the group discussion sessions, learners were seated as a group around the table. Each learner accessed an online collaborative platform through their own laptop/tablet. Video data was captured through an Intel RealSense camera in the format of .mp4 files, while the audio data was recorded as

.mp3 files through a Boya conference microphone. In total, twenty collaborative sessions, lasting from ~33min to 67min, have been analysed in this study.

3.3. Implementing the framework

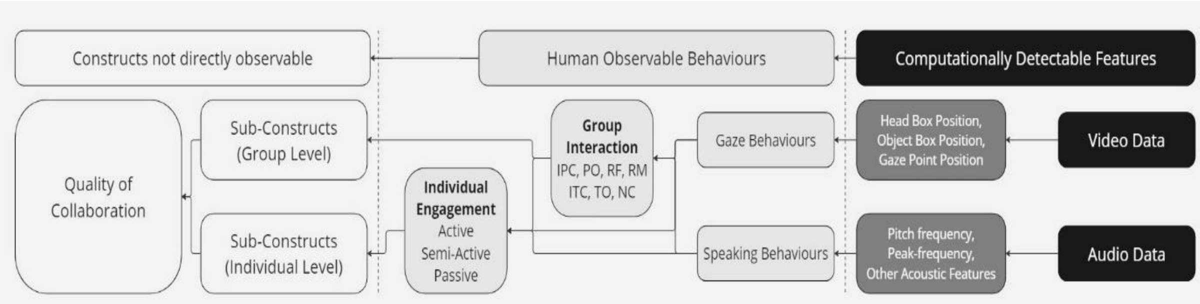

Fig. 3. An example of implementing the framework

Fig. 3. shows how the proposed framework was implemented in this study. The aim of the study is to use the MMLA system to support learners in CL by providing educational meaningful information about their learning process for them.

Computationally detectable Features. Two modalities of data were analysed in this study. A pre-trained YOLOv7 model was used to extract the head box position, object box position, and gaze position from the video data [33]. These features helped with the automatic detection of learners' gaze behaviours [23]. In terms of the audio data, Amazon Web Services (AWS) was used to automatically detect each learner's speaking behaviours through pitch frequency, peak frequency, and other acoustic features.

Human observable behaviours. For the social signals, gaze behaviours and speaking behaviours were automatically detected from the derived features presented above. Four types of gaze behaviours were defined, namely gazing at peers, gazing at laptops, gazing at tutor(s), and gazing at other objects [29]. The speaking behaviours presented whether one learner is speaking or not. Based on these social signals, a rule-based approach is used to determine seven types of group interactions, namely interacting with peers through communication (IPC), peer observation (PO), referring and following (RF), resource management (RM), interacting with the tutor(s) through communication (ITC), tutor(s) observation, and non-collaborative interactions (NC). For example, if all students were looking at their laptops/tablets, and no one was speaking, the group was inferring to have the RM interaction. If more than one student was speaking, and others were gazing at the speakers, the group was considered to have the IPC interaction. These interactions were defined based on the CL theories [17], and were illustrated to have the potential of being used to reveal the process of CL [24]. At last, each learner's individual engagement was estimated based on their social signals and the type of group interactions they were involved in. Active, Semi-active, and passive status will be used to describe their individual engagement. These statuses can be used to infer equality, individual accountability, intra-individual variability, and other sub-constructs related to the quality of CL [4].

Constructs that are not directly observable. In this study, the higher-order construct is the quality of CL. Yet, this complex concept is related to many different aspects of learning. Also, only providing an overall judgement about the quality cannot provide enough support for learners to improve their CL. Therefore, some sub-constructs were mapped to this high-order construct. From the group perspective, communication, joint information processing, coordination, and interpersonal relationships were illustrated to be important factors related to the quality of collaboration [15]. These sub-constructs will be mapped from learners' group interactions to structure the quality of CL. On the other hand, individual accountability and individual motivation will be used to assess how individuals learn and participate in the process of CL. These sub-constructs will be inferred from their engagement status.

4. Conclusion and future work

This paper presents an MMLA framework for evaluating CL from both individual and group levels. The framework is developed from the concept of "from clicks to constructs", but provides more detailed information about three different types of behavioural markers, namely social signals, group interactions, and individual engagement, which may contribute to providing more insights into the process of CL from the view of both individuals and group cohorts. Moreover, the case study was presented to give an example of how this framework can be implemented in a real-world CL context. Future work will focus more on the implementation of this framework in different contexts to explore the generalizability as well as the validity of it. Also, it would be valuable to further elaborate on the human factors and explore how this framework can help the users (learners and teachers) with their CL compared to the "black box" analysis methods and analysis pipelines with classifications of behavioural markers that are unclear, unethical, or unacceptable for their intended end-users.

5. References

[1] Roel Bosker and Tom AB Snijders. 2011. Multilevel analysis: An introduction to basic and advanced multilevel modeling. *Multilevel Analysis* (2011), 1–368.

[2] David Boud, Ruth Cohen, and Jane Sampson. 1999. Peer learning and assessment. *Assessment & evaluation in higher education* 24, 4 (1999), 413–426.

[3] Bernard P. Cohen and Elizabeth G. Cohen. 1991. From groupwork among children to R&D teams: Interdependence, interaction and productivity. *Advances in group processes* 8, (1991), 205–225.

[4] Mutlu Cukurova, Rose Luckin, Eva Millán, and Manolis Mavrikis. 2018. The NISPI framework: Analysing collaborative problem-solving from students' physical interactions. *Computers & Education* 116, (January 2018), 93–109. DOI:https://doi.org/10.1016/j.compedu.2017.08.007

[5] Daniele Di Mitri, Jan Schneider, Marcus Specht, and Hendrik Drachsler. 2018. From signals to knowledge: A conceptual model for multimodal learning analytics. *Journal of Computer Assisted Learning* 34, 4 (2018), 338–349. DOI:https://doi.org/10.1111/jcal.12288

[6] Gloria Milena Fernandez-Nieto, Vanessa Echeverria, Simon Buckingham Shum, Katerina Mangaroska, Kirsty Kitto, Evelyn Palominos, Carmen Axisa, and Roberto Martinez-Maldonado. 2021. Storytelling with learner data: Guiding student reflection on multimodal team data. *IEEE Transactions on Learning Technologies* 14, 5 (2021), 695–708.

[7] D. Randy Garrison and J. Ben Arbaugh. 2007. Researching the community of inquiry framework: Review, issues, and future directions. *The Internet and higher education* 10, 3 (2007), 157–172.

[8] Jeroen Janssen, Gijsbert Erkens, Paul A. Kirschner, and Gellof Kanselaar. 2011. Multilevel analysis in CSCL research. *Analyzing interactions in CSCL: Methods, approaches and issues* (2011), 187–205.

[9] Roger T. Johnson, David W. Johnson, and Mary B. Stanne. 1985. Effects of cooperative, competitive, and individualistic goal structures on computer-assisted instruction. *Journal of Educational Psychology* 77, 6 (1985), 668–677. DOI:https://doi.org/10.1037/0022-0663.77.6.668

[10] Marjan Laal and Seyed Mohammad Ghodsi. 2012. Benefits of collaborative learning. *Procedia - Social and Behavioral Sciences* 31, (January 2012), 486–490. DOI:https://doi.org/10.1016/j.sbspro.2011.12.091

[11] Anouschka van Leeuwen, Nikol Rummel, and Tamara van Gog. 2019. What information should CSCL teacher dashboards provide to help teachers interpret CSCL situations? *Intern. J. Comput.-Support. Collab. Learn* 14, 3 (September 2019), 261–289. DOI:https://doi.org/10.1007/s11412-019-09299-x

[12] Yingbo Ma, Mehmet Celepkolu, and Kristy Elizabeth Boyer. 2022. Detecting Impasse During Collaborative Problem Solving with Multimodal Learning Analytics. In *LAK22: 12th International Learning Analytics and Knowledge Conference* (LAK22), Association for Computing Machinery, New York, NY, USA, 45–55. DOI:https://doi.org/10.1145/3506860.3506865

[13] Jonna Malmberg, Sanna Järvelä, Jukka Holappa, Eetu Haataja, Xiaohua Huang, and Antti Siipo. 2019. Going beyond what is visible: What multichannel data can reveal about interaction in the context of collaborative learning? *Computers in Human Behavior* 96, (July 2019), 235–245. DOI:https://doi.org/10.1016/j.chb.2018.06.030

[14] Roberto Martinez-Maldonado, Dragan Gaševic, Vanessa Echeverria, Gloria Fernandez Nieto, Zachari Swiecki, and Simon Buckingham Shum. 2021. What Do You Mean by Collaboration Analytics? A Conceptual Model. *Journal of Learning Analytics* 8, 1 (2021), 126–153.

[15] Anne Meier, Hans Spada, and Nikol Rummel. 2007. A rating scheme for assessing the quality of computer-supported collaboration processes. *International Journal of Computer-Supported Collaborative Learning* 2, (2007), 63–86.

[16] A. M. O'Donnell and C. E. Hmelo-Silver. 2013. The international handbook of collaborative learning. *Introduction: What Is Collaborative Learning? an Overview* (2013), 19–40.

[17] Fan Ouyang, Weiqi Xu, and Mutlu Cukurova. 2022. An Artificial Intelligence driven Learning Analytics Method to Examine the Collaborative Problem solving Process from a Complex Adaptive Systems Perspective. *arXiv preprint arXiv:2210.16059* (2022).

[18] Theodore Panitz. 1999. Collaborative versus Cooperative Learning: A Comparison of the Two Concepts Which Will Help Us Understand the Underlying Nature of Interactive Learning. (1999).

[19] George Siemens and Ryan SJ d Baker. 2012. Learning analytics and educational data mining: towards communication and collaboration. In *Proceedings of the 2nd international conference on learning analytics and knowledge*, 252–254.

[20] Daniel Spikol, Emanuele Ruffaldi, Giacomo Dabisias, and Mutlu Cukurova. 2018. Supervised machine learning in multimodal learning analytics for estimating success in project-based learning. *Journal of Computer Assisted Learning* 34, 4 (2018), 366–377. DOI:https://doi.org/10.1111/jcal.12263

[21] Daniel Spikol, Emanuele Ruffaldi, Lorenzo Landolfi, and Mutlu Cukurova. 2017. Estimation of Success in Collaborative Learning Based on Multimodal Learning Analytics Features. In *2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT)*, 269–273. DOI:https://doi.org/10.1109/ICALT.2017.122

[22] Mark Summers and Simone Volet. 2010. Group work does not necessarily equal collaborative learning: evidence from observations and self-reports. *Eur J Psychol Educ* 25, 4 (December 2010), 473–492. DOI:https://doi.org/10.1007/s10212-010-0026-5

[23] Qi Zhou, Amartya Bhattacharya, Wannapon Suraworachet, Hajime Nagahara, and Mutlu Cukurova. 2023. Automated Detection of Students' Gaze Interactions in Collaborative learning Videos: A Novel Approach. In *EC-TEL 2023, Aveiro, Potugal, September 04-08, 2023*

[24] Susan R. Swing and Penelope L. Peterson. 1982. The Relationship of Student Ability and Small-Group Interaction to Student Achievement. *American Educational Research Journal* 19, 2 (January 1982), 259–274. DOI:https://doi.org/10.3102/00028312019002259

[25] Alessandro Vinciarelli, Maja Pantic, and Hervé Bourlard. 2009. Social signal processing: Survey of an emerging domain. *Image and Vision Computing* 27, 12 (November 2009), 1743–1759. DOI:https://doi.org/10.1016/j.imavis.2008.11.007

[26] Noreen M. Webb. 1980. An Analysis of Group Interaction and Mathematical Errors in Heterogeneous Ability Groups. *British Journal of Educational Psychology* 50, 3 (1980), 266–276. DOI:https://doi.org/10.1111/j.2044-8279.1980.tb00810.x

[27] Noreen M. Webb. 1993. Collaborative group versus individual assessment in mathematics: Processes and outcomes. *Educational Assessment* 1, 2 (1993), 131–152.

[28] Alyssa Friend Wise, Simon Knight, and Simon Buckingham Shum. 2021. Collaborative learning analytics. In *International handbook of computer-supported collaborative learning*. Springer, 425–443.

[29] Qi Zhou, Wannapon Suraworachet, Oya Celiktutan, and Mutlu Cukurova. 2022. What Does Shared Understanding in Students' Face-to-Face Collaborative Learning Gaze Behaviours "Look Like"? In *Artificial Intelligence in Education* (Lecture Notes in Computer Science), Springer International Publishing, Cham, 588–593. DOI:https://doi.org/10.1007/978-3-031-11644-5_53

[30] Qi Zhou, Wannapon Suraworachet, and Mutlu Cukurova. 2021. Different modality, different design, different results: Exploring self-regulated learner clusters' engagement behaviours at individual, group and cohort activities. In *CEUR Workshop Proceedings*, CEUR, 28–40.

[31] Qi Zhou, Wannapon Suraworachet, Stanislav Pozdniakov, Roberto Martinez-Maldonado, Tom Bartindale, Peter Chen, Dan Richardson, and Mutlu Cukurova. 2021. Investigating Students' Experiences with Collaboration Analytics for Remote Group Meetings. In *Artificial Intelligence in Education* (Lecture Notes in Computer Science), Springer International Publishing, Cham, 472–485. DOI:https://doi.org/10.1007/978-3-030-78292-4_38

[32] Mutlu Cukurova, Qi Zhou, Daniel Spikol, Lorenzo Landolfi, 2020. Modelling collaborative problem-solving competence with transparent learning analytics: is video data enough?. *In Proceedings of the tenth international conference on learning analytics & knowledge, pp. 270-275*
[33] CrowdHuman Dataset, https://www.crowdhuman.org/, last accessed 2023/01/16.