Exploring Indicators for Collaboration Quality and its Dimensions in Classroom Settings using Multimodal Learning Analytics

Pankaj Chejara[1], Luis P. Prieto[2], María Jesús Rodríguez-Triana[1], Adolfo Ruiz-Calleja[1], Reet Kasepalu[1], Irene-Angelica Chounta[3], and Bertrand Schneider[4]

[1] Tallinn University, Tallinn, Estonia
{pankajch,mjrt,adolfo,reetkase}@tlu.ee
[2] Universidad de Valladolid, Valladolid, Spain
luispablo.prieto@uva.es
[3] University of Duisburg-Essen, Duisburg, Germany
irene-angelica.chounta@uni-due.de
[4] Harvard University, Cambridge, MA, USA
bertrand_schneider@g.harvard.edu

Abstract. Multimodal Learning Analytics researchers have explored relationships between collaboration quality and multimodal data. However, the current state-of-art research works have scarcely investigated authentic settings and seldom used video data that can offer rich behavioral information. In this paper, we present our findings on potential indicators for collaboration quality and its underlying dimensions such as argumentation, and mutual understanding. We collected multimodal data (namely, video and logs) from 4 Estonian classrooms during authentic computer-supported collaborative learning activities. Our results show that vertical head movement (looking up and down) and mouth region features could be used as potential indicators for collaboration quality and its aforementioned dimensions. Also, our results from clustering provide indications of the potential of video data for identifying different levels of collaboration quality (e.g., high, low, medium). The findings have implications for building collaboration quality monitoring and guiding systems for authentic classroom settings.

Keywords: Multimodal Learning Analytics · MMLA · Computer-Supported Collaborative Learning · CSCL · Collaboration Quality · Correlation Analysis · Machine Learning · Clustering · Facial Action Units.

1 Introduction

Teachers are often expected to orchestrate Computer-Supported Collaborative Learning (CSCL) activities, monitor groups, identify groups who need help, and then offer support. These tasks are difficult for teachers to carry out efficiently in the classroom given their limited time [32], and especially when the number of groups increases [5]. Moreover, the COVID-19 pandemic pushed teachers to design and enact online group activities, which further complicated the aforementioned tasks of orchestrating and monitoring [20]. Since then, the organization of CSCL activities in hybrid settings has become mainstream, raising the need for supporting teachers during those activities.

To reach that goal, researchers have analyzed digital log data from the group's activities to gain insight into collaboration behavior and to develop monitoring systems [5]. However, this log-based analysis offers only a partial understanding of CSCL in classroom settings involving physical and digital spaces of interaction. The need for a more holistic understanding of collaborative learning (& learning in general) has led to the use of additional data sources [24, 10]. This field of research is known as Multimodal Learning Analytics (MMLA) [1, 17].

To better understand collaboration behavior, MMLA Researchers have used different types of data such as audio, video, eye-gaze, skin-conductance, and body-pose [18, 24]. These data have been used with various analysis methods ranging from statistical analysis [2], to unsupervised [13], and supervised machine learning [33, 30]. With these methods, MMLA researchers have identified several data features as potential indicators for collaboration behavior [18, 24], and shown the feasibility of building automated models for collaboration estimation [33, 3, 19]. For example, speaking time distribution [18] and physiological synchrony (Directional Agreement) were found as potential indicators for overall collaboration quality [23].

The current state of MMLA research has enabled the development of mirroring systems [29] which provides visualization of the group's activities and their overall quality of collaboration [16]. However, such systems are limited in actionability (i.e., lack of suggestions on potential intervention strategies). Furthermore, such systems do not offer any information on the resultant estimation of collaboration quality which often leaves teachers clueless about what could be done to improve the current situation. Consequently, there is a need to develop guiding systems that can offer actionable feedback to teachers which would require knowledge about underlying dimensions of collaboration quality (refer for more details, Collaboration Intervention Model from [15]).

Certain research studies have explored collaboration quality dimensions using log [5] and audio [19] data. However, these studies along with a majority of current research studies are conducted in controlled settings [24] which allows higher quality data. On the other hand, there are inherent data noise issues when working within *authentic* classroom settings [28]. For example, the use of audio data in such settings is most likely to result in lower-quality data due to background noise. Consequently, there is a need to explore alternate data sources which could be employed in authentic settings for understanding and estimating collaboration quality dimensions.

To address the aforementioned gaps of the lack of research on the exploration of the relationship between *multimodal data and collaboration quality dimensions* in *authentic classroom settings*, we set up the following research questions: **(RQ1)** What is the relationship between video-log data and collaboration quality (and its dimensions as per Rummel et al., 2011 [22]) of small groups in authentic classroom settings? **(RQ2)**; To what extent can video-log data features reveal differences in terms of collaboration quality and its dimensions in authentic classroom settings? This paper is structured as follows. Section 2 provides the state-of-the-art MMLA research in CSCL. We then explain the study context, the data collection tool used, the procedure followed to conduct the study, and data features extracted from the data in Section 3. Section 4 presents our methods for annotation and analysis, followed by the results and their implications in Sections 5 and 6, respectively. Finally, we conclude in Section 7 with future research work directions.

2 Related work

The use of MMLA for understanding and modeling collaboration settings has grown significantly in the past decade [24, 18, 6]. Researchers have shown the potential of multimodal data such as audio, video, eye gaze, and heart rate, towards unraveling the complexity of collaboration [30, 23, 19, 3]. These data have been used in current state-of-the-art research for extracting features, ranging from speaking time [33], turn-taking [3], speech features (e.g., coh-matrix) [21] to facial action units [2, 26], emotions [12], distance between hands [30]. Using these features, researchers have explored the relationship between multimodal data and collaboration measures (e.g., collaboration quality [13] or collaborative experience [26]). Several data features have been identified as potential indicators for collaboration, e.g., equal speaking time distribution [18], number of words spoken [21], the distance between participants [30].

There are a few research studies that have even delved deeper and explored collaboration dimensions [19, 4]. For example, Pugh et al., [19] showed the potential of language models to detect aspects of collaborative problem solving, i.e., knowledge exchange, and coordination. Similarly, Cai et al., [2] explored the relationship between facial action units and different states of learning as per the ICAP framework during a group-work activity. Hayashi et al., [12] explored relationship between emotions and collaboration quality dimensions. These research studies provide preliminary evidence of the potential of multimodal data to identify underlying dimensions of collaboration.

We identify two research gaps in current MMLA research on understanding collaboration behavior. The first gap is that researchers have seldom investigated authentic classroom settings for understanding collaboration or exploring the relationship between multimodal data and collaboration [6, 24]. Consequently, the applicability of knowledge gained (or automated models developed) from research studies conducted in laboratory settings to authentic classroom settings is still in question. The second gap is that the current research focused primarily on high-level collaboration measures (e.g., collaboration quality [33], artifact quality [30]), and despite the rich behavioral and interaction data captured through video, current state-of-art has scarcely used it. Thus, there is a

Table 1: Dataset characteristics

Class	Language	Subject	Groups	Students	Data instances (30s windows)
10th	Estonian	Office work	3	10	193
10th	English	Biology	7	25	88
9th	Estonian	Homeroom teacher lesson	7	24	273
12th	English	Biology	3	10	227

lack of knowledge about the relationship between video data and dimensions of collaboration quality.

3 Study setup

This section describes the study context, data collection setup & procedure, and data features.

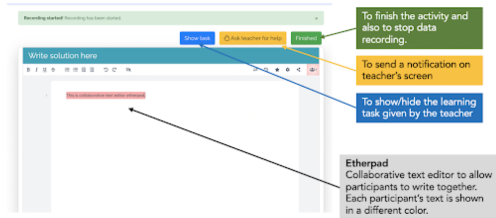
3.1 Context

The study was conducted in Estonian upper secondary school classrooms with four different teachers during biology, office work, and homeroom teacher lessons. The students were mainly of Estonian background. The languages of communication were Estonian and English. There were a total of 69 students in 20 groups. We had to discard the data from 7 groups due to missing video recordings from those groups. Table 1 provides details about the study contexts.

3.2 Activity tasks

While the specific topic of the activities varied depending on the subject, each lesson had the same structure. This required students to discuss a certain topic face-to-face and write their outcomes in a collaborative text editor (e.g., Etherpad). In the homeroom teacher's lesson, the students were given the task of planning a class trip involving collaboratively selecting a destination, allocating a budget (travel, meals, accommodation), and creating a schedule for the entire trip. In the office work session, the task involved discussing and synthesizing the process of archiving (e.g., the preparation of the description and the necessity for this, and who the end-user of an archive's description is). In the biology session, the groups were given worksheets that had questions on DNA sequencing and mutation (e.g., on the effect of GTA $\rightarrow$ GTT mutation) to be answered in the collaborative text editor. Figure 1 shows an example of the student arrangement and the collaborative editor used in the activities.

(a)



(b)

Fig. 1: (a) Students working on the collaborative activity in the classroom (b) Collaborative activity space in CoTrack

3.3 Data collection tool

To carry out the study we used CoTrack[5] [3]. This web-based application offers a collaborative text editor (Etherpad[6]) for students to work on a given task (Figure 1b). In addition, CoTrack records audio/video using computer's microphone/camera along with log data and process these data in real-time to extract data features (e.g., the number of characters added by a student in the group). These data features are used to generate a dashboard to help teachers monitor the groups' work.

3.4 Procedure

Prior to the research study, a researcher from education science (also a co-author of this paper) co-designed the lesson with the teacher who was to instruct the collaborative activity. The consent was taken from the students in advance of the study (parents' consent was taken in the case of students younger than 18 years). On the day of enactment, the same researcher was also present in the classroom and gave a brief introduction about the research to the students. Following that, the teacher assigned groups to the students who then used CoTrack to work on the given activity collaboratively. The average duration of the activity was approximately 30 minutes.

3.5 Features

We extracted features based on the current state-of-the-art in MMLA for collaborative learning. From the video, we decided to use facial action units because of preliminary evidence of a relationship between action units and collaboration quality [2, 7]. We also decided to use head pose as a proxy of eye gaze which has

[5] https://www.cotrack.website
[6] https://www.etherpad.org

been found as a good indicator for collaboration quality behavior [18]. Previous research has found that head pose contributes 68% towards eye-gaze direction [31]. In addition, we also extracted features that could be used as a proxy for speaking, a widely used feature in MMLA [33]. Due to the limitations of captured video (i.e., recording mainly above shoulder area), we decided not to extract features, e.g., body posture, hand-movement, and interaction with artefact. We did not use audio due to background noise affecting the audio quality, a frequently reported issue with MMLA research in the classroom [8]. Instead, we used video-based features for the speaking activity. Such vision-based speaking activity features could be useful in authentic classroom settings where audio data is often of a lower quality due to background audio noise. We used mouth region as a proxy for speaking based on computer vision research work focusing on speaking activity detection using image processing [27]. In their research work, the mouth area region (the grey area inside a bounding box around the mouth of the speaker) has been used to distinguish speaking activity from non-speaking activity.

From logs, we extracted simple features, namely the number of characters written or deleted. Our decision was based on research that highlighted individual participation as one of the quantitative key features for collaborative learning [34]. Furthermore, we also found in our previous studies that models based on these log features in addition to aforementioned audio features (i.e., turn-taking, speaking time) performed well on collaboration quality estimation tasks [4, 3].

4 Methods

4.1 Annotation

We used Rummel et al., 2011 rating scheme to annotate the collaboration quality [22]. This rating scheme specifies seven dimensions of collaboration quality: argumentation, sustaining mutual understanding, cooperative orientation, knowledge exchange, collaboration flow, individual task orientation, as well as structuring problem solving and time management. We annotated these seven dimensions for every 30 seconds time window, as recommended in prior works [3]. Each dimension was assigned a score on a 5-Likert scale [-2 to +2] and added together to compute the overall collaboration quality score. The rating scheme was adapted to the set of activities conducted in this study. The first 10 minutes of video of a group was coded independently by two researchers and conflict among their scores was discussed. The rating scheme was used to train four MA students who went through two rounds of annotation (one in a group, another independent). After the training, all video recordings were assigned to them for annotation. The inter-rater reliability (Cohen's Kappa = .61) was above a substantial level for all the dimensions

Table 2: Correlation between multimodal features and collaboration quality (STR: structuring problem solving process and time management, ITO: individual task orientation, CF: collaboration flow, SMU: sustaining mutual understanding, ARG: argumentation, CO: cooperative orientation, KE: knowledge exchange, CQ: collaboration quality, avg: average, sd: standard deviation).

Feature	**Aggregation** window	group	**STR**	**ITO**	**CF**	**SMU**	**ARG**	**CO**	**KE**	**CQ**
AU01	count	avg	-.38	—	—	—	—	—	—	—
AU01	count	sd	-.35	—	—	—	—	—	—	—
Head rotate x-axis	std	avg	—	—	.30	.30	—	—	—	.31
Head rotate y-axis	std	avg	—	.32	—	—	—	—	—	—
Mouth_area	avg	avg	—	.34	.30	—	.31	—	—	.30
Chars add	sum	avg	—	.31	—	—	—	—	—	—
Chars del	sum	avg	—	.31	—	—	—	—	—	—

4.2 Analysis

We performed a Spearman correlation analysis to identify the relationship between multimodal data and collaboration quality dimensions. We used this non-parametric test because the underlying data was not following normal distribution as found in the normality test (Shapiro-Wilk). Next, we used the K-means clustering technique to investigate whether the used features could reveal differences among collaboration quality scores. For clustering, we used the correlated features. To identify the most important features in clustering, we followed the approach proposed in [14]: first, we considered cluster membership as class labels and then trained a random forest model using our features and labels; next, we identified the most important features using the trained random forest model. We used the identified features to interpret the resultant clusters. Finally, we checked the differences in resultant clusters' collaboration quality scores for statistical significance and for that, we used the Kruskal-Wallis test.

5 Results

Relationship between video-log features and collaboration quality (and its dimensions). Table 2 reports statistically significant (p-value<.001) correlation measures between video-log features and collaboration quality (and its dimensions) scores. We found that the average frequency of facial Action Unit 1 (AU01), the inner brow raiser, was negatively correlated (ρ=-.38) with the structuring problem solving and time management (STR) dimension. We also observed a similar relationship when the standard deviation of AU01 was taken at the group level. That means that as the variation of group members' frequency of inner brow raiser goes higher, a lower rating of STR was observed. In addition to this, we found that variation in vertical head movement[7] was positively cor-

[7] Head rotation along the x-axis (i.e., moving the head up and down)

related with collaboration quality (ρ=.31), collaboration flow (ρ=.30), and sustaining mutual understanding (ρ=.30). We can also notice from the table that variation in horizontal head movement[8] was found to be positively correlated with individual task orientation (ρ=.32). The average mouth region area at the group level was found to be positively correlated with argumentation (ρ=.31), individual task orientation (ρ=.34), collaboration flow (ρ=.30), and collaboration quality (ρ=.30). We found a positive correlation (ρ=.31) between individual task orientation and Etherpad activities (i.e., the number of characters added or deleted).

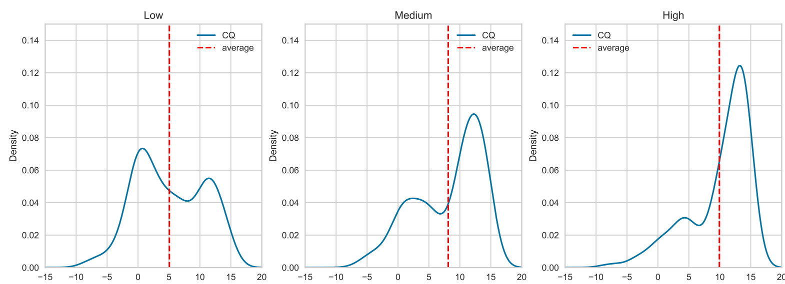


Fig. 2: K-means clusters and distribution of collaboration quality scores of those clusters

Unsupervised machine learning revealing differences amongst students' group collaboration quality using multimodal data. Our results from K-means clustering showed that the resulting clusters had differences in their collaboration quality scores. Figure 2 shows the distribution of collaboration quality scores for each cluster. When comparing with the other two clusters, the first cluster had lower scores for collaboration quality (average CQ score = 5). Therefore, we interpreted the cluster as having low collaboration quality. The third cluster had the highest scores for collaboration quality (average CQ score = 9.9) among the three, thus, interpreted as having high collaboration quality. The second cluster had intermediate collaboration quality scores (average CQ score = 8.1), therefore we considered it as having medium collaboration quality. We also analyzed clusters obtained from k-means with respect to the scores of collaboration quality dimensions and observed a similar pattern of high, medium, and low collaboration quality. Figure 3 shows the distribution of the top five most important features for resultants clusters. These features were the average frequency of lip presser (AU24), lips apart (AU25) action units, variation among lip presser action unit at group level, vertical head movement, and mouth area region. The cluster with relatively high collaboration quality scores (↑) was found to have a higher average frequency of lips apart (↑), lower average frequency of

[8] Head rotation along the y-axis (i.e., moving the head left and right)

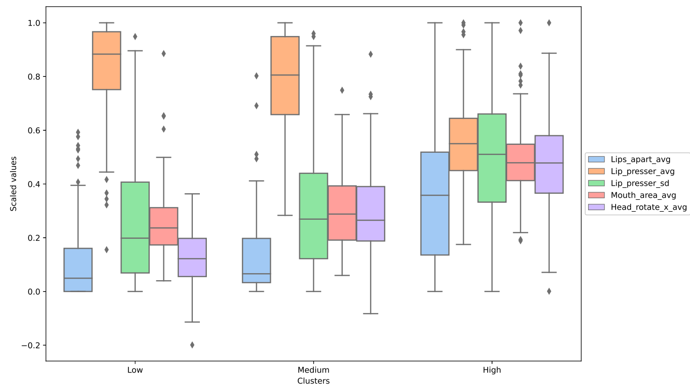


Fig. 3: Cluster-wise distribution of top five important features

lip presser (↓), higher variation in the frequency of lip presser (↑), and higher values for vertical head movement (↑) and mouth area region (↑).

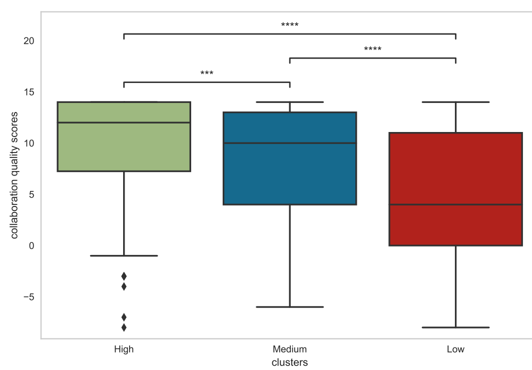


Fig. 4: Post-hoc analysis results (***: p-value $\leq$ 1.00e-03; ****: p-value $\leq$ 1.00e-04)

Statistical significance of differences in collaboration quality scores of resultant clusters. Our results from Kruskal-Wallis test showed that the differences between the three clusters in terms of collaboration quality and its dimensions were of statistical significance (P-value < .001). The posthoc results from Dunn's test also revealed that the pair-wise differences among clusters were statistically significant (Figure 4).

6 Discussion

6.1 What is the relationship between multimodal data collected and collaboration quality (and its dimensions) of small groups in authentic classroom settings? (RQ1)

Finding 1: *Inner brow raiser (AU01) is negatively correlated with structuring problem solving and time management dimension of collaboration quality.*
The structuring problem-solving dimension involves deciding the strategy for solving the given problem in the group. This might trigger negative emotions (e.g., frustration or confusion) in the group if there are group participants who disagree with the decision. These negative emotions could also be due to the pressure of completing the activity within the given time. Prior research suggested a relationship between inner brow raiser (AU01) and frustration [7]. This might indicate that frustration may appear when groups cannot structure their collaborative activity. This interpretation aligns with previous research on the investigation of the relationship between emotions and dimensions of collaboration quality [12].

Finding 2: *Vertical head movement is positively correlated with collaboration quality, collaboration flow, and sustaining mutual understanding.*
The relationship between higher vertical head movement (i.e., looking up and down) in the group and collaboration quality dimensions could be either due to participants looking at the screen or looking at other group members while discussing. The latter may be an indication of participation. The same feature also correlated with collaboration flow and sustaining mutual understanding. A vertical head movement could also be due to head nods which are found as non-verbal cues for understanding [11]. However, the collaborative activities described in this paper involved the use of Etherpad (a collaborative text editor) which may have caused participants to look at the screen frequently.

Finding 3: *Horizontal head movement is positively correlated with individual task orientation.*
The horizontal head movement (i.e., looking left and right) positively correlated with the individual task orientation dimension. This dimension refers to participants' motivation, externalized as active participation in the activity, and consequently can be measured in terms of participants' contribution towards the problem solution. As the participants were interacting with Etherpad, this might have resulted in looking at the screen. While doing that their head movement in horizontal direction can be explained by looking across the screen and also at others' contributions in Etherpad. Another explanation could be that horizontal head movement may have been caused by participants looking at their group members sitting on their left or right side, depending on their sitting arrangement.

Finding 4: *Mouth area region positively correlates with collaboration quality, argumentation, collaboration flow, and individual task orientation.*
We used this feature (mouth area region) as a proxy for speaking because the increased average value of mouth region pixels can be used as a visual cue for

speaking activity [27]. Previous research pointed to speaking time as an indicator of collaboration [21]. Our results confirm that speaking is positively associated with collaboration quality and its underlying dimensions.

Finding 5: *The number of characters written and deleted is positively correlated with individual task orientation dimensions.*
The individual task orientation dimension looks upon individual participation in solving the given problem. This participation could be partially measured by the number of writing actions (e.g., writing or deleting text). This may explain the positive relationship between individual task orientation and quantitative log features.

6.2 To what extent, video-log data features can reveal differences in terms of collaboration quality and its dimensions in authentic classroom settings? (RQ2)

Finding 6: *K-means clusters of video-log data match three different levels of collaboration quality (high, medium, and low).*
The emerging clusters demonstrated a different distribution of scores of collaboration quality and could be characterized as of high, medium, and low collaboration quality. The high collaboration quality cluster had comparatively higher values for mouth region area, vertical head movement, lips apart action unit, variation in the lip presser action unit, and lower values for lip presser action unit. For low collaboration quality clusters, these values were inverted. The higher value of lips apart and the lower value of lip presser action units in the high collaboration quality cluster could be explained by both negative and positive emotions occurring during group activity [26, 12]. Lip presser is often related to the expression of anger [25]. Thus, a higher occurrence of lip presser at the group level may indicate a higher level of anger. However, the higher variation in lip presser action units was observed to have higher collaboration quality scores. This could be explained by the conflicts which usually arise in the group during the activity and which are likely to trigger anger. This situation is unlikely to be resolved if all participants experience the same level of anger. To be resolved, some participants – possibly the ones in a relaxed state – need to mediate the situation. This could explain higher variation in lip presser in the cluster of high collaboration quality scores. The high value of the lips apart action unit could be explained by prior research [2] suggesting that AU25 was observed more frequently in the constructive state than the interactive and passive state (as per ICAP framework). Vertical head movement could be due to students' focusing on Etherpad and their peers causing the up-down head movement. Moreover, the head nods used for acknowledgment and agreement might also cause the up-down head movement. The high value on the mouth area region could indicate more speaking activity in the group. Speaking has previously been found as a potential indicator for collaboration quality [18].

Finding 7: *The scores of the underlying dimension of collaboration quality in resultant K-means clusters also have statistically significant differences.*
Our findings suggest that there were statistically significant differences among

clusters in terms of their scores of collaboration quality dimensions. These results suggest that video data along with log data could be used in authentic classroom settings for collaboration quality (and its dimensions) estimation tasks. For example, vertical head movement during CSCL activities in authentic classroom settings is found to contribute towards differentiating three different levels of collaboration quality scores. Prior research has shown the potential of log data [5] and audio data [4, 19] for classifying collaboration quality dimensions. With our findings, we extend those research works and present preliminary evidence over the potential use of video data along with logs for estimating collaboration quality and its dimensions.

6.3 Limitations

The presented study has three main limitations. The first limitation is related to the seating arrangement of students. Due to the nature of the activities and the availability of resources in the school, one session was conducted in a computer lab where participants were sitting side by side. The head movements along the y-axis (looking left and right) may have been caused by participants looking at their group members sitting on their left and right sides. Thus, the found relationship between a head movement along the y-axis and individual task orientation dimension needs further validation. The second limitation is the limited scope of the findings' generalizability. As the participants in the study were upper secondary students mostly of Estonian background, it restricts us from making any claim over the applicability of the research findings on students from a different age or cultural background which may condition their communication and collaboration style. The third limitation is that this paper mainly analyzed all the multimodal features independently and did not utilize interaction between those features in the exploration.

7 Conclusion & Future work

This paper addresses the current gap in our understanding of the relationship between multimodal (video and log) data and collaboration quality dimensions in authentic classroom settings. We collected data from Estonian classroom CSCL activities and performed correlation analysis to identify relationships between features extracted from data and collaboration quality dimensions. Our results showed that simple features from the video and log data, more concretely inner brow raiser (AU01), mouth region area, vertical head movement (i.e., up-down head movement), and number of characters written could be used in authentic settings as indicators of collaboration quality dimensions (namely structuring problem solving and time management, collaboration flow, argumentation and sustaining mutual understanding). Among these indicators, the features capturing vertical head movement and mouth region area could potentially be used as an overall indicator of collaboration quality in authentic classroom settings. We also provided preliminary evidence on the potential of video data along with the logs in estimating collaboration quality (and its dimensions) using unsupervised machine learning. Future research should verify whether these findings apply to other contexts and can be generalized.

The current MMLA research in CSCL often employs supervised machine learning techniques which require a large amount of data to work with. However, given the need for human resources for data annotation, the final datasets are often of a smaller size. The use of unsupervised learning can help in dealing with the issue of the lack of a large amount of annotated data. With the use of unsupervised machine learning, we foresee a hybrid use of supervised and unsupervised learning techniques to model collaboration behavior with multimodal data also suggested in the prior work [35]. In our future work, we plan to assess the potential of video data for the prediction of collaboration quality and its dimensions utilizing both supervised and unsupervised learning techniques. We will also analyze multiple features together to understand the found relationship better following prior work [9]. We plan to look at the differences in the relationship between multimodal data and collaboration dimensions across a wider range of learning activities, student groups, and education levels.

8 Acknowledgements

The presented work has been partially funded by the Estonian Research Council's Personal Research Grant (PRG) under grant number PRG1634. It also has been supported by grant RYC2021-032273-I, financed by MCIN/ AEI/ 10.13039/501100011033 and the European Union's "NextGenerationEU/PRTR".

References

1. Blikstein, P., Worsley, M.: Multimodal learning analytics and education data mining: using computational technologies to measure complex learning tasks. Journal of Learning Analytics **3**(2), 220–238 (Sep 2016)
2. Cai, Y., Shimojo, S., Hayashi, Y.: Observing facial muscles to estimate the learning state during collaborative learning: A focus on the ICAP framework. ICCE 2020 - 28th International Conference on Computers in Education, Proceedings **1**, 119–126 (2020)
3. Chejara, P., Prieto, L.P., Rodriguez-Triana, M.J., Kasepalu, R., Ruiz-Calleja, A., Shankar, S.K.: How to build more generalizable models for collaboration quality? lessons learned from exploring multi-context audio-log datasets using multimodal learning analytics. In: LAK23: 13th International Learning Analytics and Knowledge Conference. p. 111–121. LAK2023, ACM, NY, USA (2023)
4. Chejara, P., Prieto, L.P., Ruiz-Calleja, A., Rodríguez-Triana, M.J., Shankar, S.K., Kasepalu, R.: Efar-mmla: An evaluation framework to assess and report generalizability of machine learning models in mmla. Sensors **21**(8), 2863 (Apr 2021)
5. Chounta, I., Avouris, N.M.: Towards a time series approach for the classification and evaluation of collaborative activities. Comput. Informatics **34**(3), 588–614 (2015)

6. Chua, Y.H.V., Dauwels, J., Tan, S.C.: Technologies for automated analysis of co-located, real-life, physical learning spaces: Where are we now? In: LAK19: 9th International Conference on Learning Analytics & Knowledge. p. 11–20. LAK19, ACM, NY, USA (2019)
7. Craig, S.D., D'Mello, S., Witherspoon, A., Graesser, A.: Emote aloud during learning with AutoTutor: Applying the Facial Action Coding System to cognitive - Affective states during learning. Cognition and Emotion **22**(5), 777–788 (2008)
8. Crescenzi-Lanna, L.: Multimodal Learning Analytics research with young children: A systematic review. British Journal of Educational Technology **51**(5), 1485–1504 (2020)
9. Cukurova, M., Luckin, R., Milln, E., Mavrikis, M.: The nispi framework. Comput. Educ. **116**(C), 93–109 (jan 2018), https://doi.org/10.1016/j.compedu.2017.08.007
10. Di Mitri, D., Schneider, J., Specht, M., Drachsler, H.: From signals to knowledge: A conceptual model for multimodal learning analytics. Journal of Computer Assisted Learning **34**(4), 338–349 (2018)
11. Duncan, S.: Some signals and rules for taking speaking turns in conversations. Journal of personality and social psychology **23**(2), 283 (1972)
12. Hayashi, Y.: Detecting collaborative learning through emotions: An investigation using facial expression recognition. In: Intelligent Tutoring Systems: 15th International Conference, ITS 2019, Kingston, Jamaica, June 3–7, 2019, Proceedings 15. pp. 89–98. Springer (2019)
13. Huang, K., Bryant, T., Schneider, B.: Identifying collaborative learning states using unsupervised machine learning on eye-tracking, physiological and motion sensor data. In: EDM 2019 - Proceedings of the 12th International Conference on Educational Data Mining. pp. 318–323. EDM (2019)
14. Ismaili, O.A., Lemaire, V., Cornuéjols, A.: A supervised methodology to measure the variables contribution to a clustering. In: Neural Information Processing: 21st International Conference, ICONIP 2014, Kuching, Malaysia, November 3-6, 2014. Proceedings, Part I 21. pp. 159–166. Springer (2014)
15. Kasepalu, R., Prieto, L.P., Ley, T., Chejara, P.: Teacher artificial intelligence-supported pedagogical actions in collaborative learning coregulation: A wizard-of-oz study. Frontiers in Education **7** (2022)
16. Martinez-Maldonado, R., Clayphan, A., Yacef, K., Kay, J.: Mtfeedback: providing notifications to enhance teacher awareness of small group work in the classroom. IEEE Transactions on Learning Technologies **8**(2), 187–200 (2014)
17. Ochoa, X., Worsley, M.: Augmenting learning analytics with multimodal sensory data. Journal of Learning Analytics **3**(2), 213–219 (2016)
18. Praharaj, S., Scheffel, M., Drachsler, H., Specht, M.: Literature review on co-located collaboration modeling using multimodal learning analytics—can we go the whole nine yards? IEEE Transactions on Learning Technologies **14**(3), 367–385 (2021)
19. Pugh, S.L., Rao, A., Stewart, A.E., D'Mello, S.K.: Do Speech-Based Collaboration Analytics Generalize Across Task Contexts? In: LAK22: 12th International Learning Analytics and Knowledge Conference. pp. 208–218. ACM, Online, USA (mar 2022)
20. Rannastu-Avalos, M., Siiman, L.A.: Challenges for distance learning and online collaboration in the time of covid-19: Interviews with science teachers. In: Collaboration Technologies and Social Computing: 26th International Conference, CollabTech 2020, Tartu, Estonia, September 8–11, 2020, Proceedings 26. pp. 128–142. Springer (2020)

21. Reilly, J.M., Schneider, B.: Predicting the quality of collaborative problem solving through linguistic analysis of discourse. In: Desmarais, M.C., Lynch, C.F., Merceron, A., Nkambou, R. (eds.) Proceedings of the 12th International Conference on Educational Data Mining, EDM 2019. pp. 149–157. International Educational Data Mining Society (IEDMS), Montréal, Canada (jul 2019)
22. Rummel, N., Deiglmayr, A., Spada, H., Kahrimanis, G., Avouris, N.: Analyzing collaborative interactions across domains and settings: An adaptable rating scheme, pp. 367–390. Springer US, Boston, MA (2011)
23. Schneider, B., Dich, Y., Radu, I.: Unpacking the relationship between existing and new measures of physiological synchrony and collaborative learning: A mixed methods study. International Journal of Computer-Supported Collaborative Learning **15**, 89–113 (2020)
24. Schneider, B., Sung, G., Chng, E., Yang, S.: How Can High-Frequency Sensors Capture Collaboration? A Review of the Empirical Links between Multimodal Metrics and Collaborative Constructs. Sensors **21** (2021)
25. Sell, A., Cosmides, L., Tooby, J.: The human anger face evolved to enhance cues of strength. Evolution and Human Behavior **35**(5), 425–429 (2014)
26. Sharma, K., Papavlasopoulou, S., Giannakos, M.: Joint emotional state of children and perceived collaborative experience in coding activities. In: Proceedings of the 18th ACM International Conference on Interaction Design and Children, IDC 2019. pp. 133–145 (2019). https://doi.org/10.1145/3311927.3323145
27. Siatras, S., Nikolaidis, N., Krinidis, M., Pitas, I.: Visual lip activity detection and speaker detection using mouth region intensities. IEEE Transactions on Circuits and Systems for Video Technology **19**(1), 133–137 (2008)
28. Smith, J., Bratt, H., Richey, C., Bassiou, N., Shriberg, E., Tsiartas, A., D'Angelo, C., Alozie, N.: Spoken interaction modeling for automatic assessment of collaborative learning. In: Proc. Speech Prosody 2016. pp. 277–281 (2016)
29. Soller, A., Martínez, A., Jermann, P., Muehlenbrock, M.: From mirroring to guiding: A review of state of the art technology for supporting collaborative learning. Int. J. Artif. Intell. Ed. **15**(4), 261–290 (dec 2005)
30. Spikol, D., Ruffaldi, E., Dabisias, G., Cukurova, M.: Supervised machine learning in multimodal learning analytics for estimating success in project-based learning. Journal of Computer Assisted Learning **34**(4), 366–377 (2018)
31. Stiefelhagen, R., Zhu, J.: Head orientation and gaze direction in meetings. In: CHI '02 Extended Abstracts on Human Factors in Computing Systems. pp. 858–859. No. 1 in CHI EA '02, ACM, New York, NY, USA (2002)
32. Van Leeuwen, A., Rummel, N., Van Gog, T.: What information should CSCL teacher dashboards provide to help teachers interpret CSCL situations? International Journal of Computer-Supported Collaborative Learning **14**(3), 261–289 (2019)
33. Viswanathan, S.A., VanLehn, K.: Using the tablet gestures and speech of pairs of students to classify their collaboration. IEEE Trans. Learn. Technol. **11**(2), 230–242 (2018)
34. Weinberger, A., Fischer, F.: A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. Computers & education **46**(1), 71–95 (2006)
35. Worsley, M.: (dis)engagement matters: Identifying efficacious learning practices with multimodal learning analytics. In: LAK18: 8th International Conference on Learning Analytics and Knowledge. p. 365–369. LAK '18, ACM, NY, USA (2018)