Education and Information Technologies
https://doi.org/10.1007/s10639-023-12315-1

# Detecting non-verbal speech and gaze behaviours with multimodal data and computer vision to interpret effective collaborative learning interactions

**Qi Zhou**[1] · **Wannapon Suraworachet**[1] · **Mutlu Cukurova**[1]

Received: 16 March 2023 / Accepted: 27 October 2023


**Abstract**
Collaboration is argued to be an important skill, not only in schools and higher education contexts but also in the workspace and other aspects of life. However, simply asking students to work together as a group on a task does not guarantee success in collaboration. Effective collaborative learning requires meaningful interactions among individuals in a group. Recent advances in multimodal data collection tools and AI provide unique opportunities to analyze, model and support these interactions. This study proposes an original method to identify group interactions in real-world collaborative learning activities and investigates the variations in interactions of groups with different collaborative learning outcomes. The study was conducted in a 10-week long post-graduate course involving 34 students with data collected from groups' weekly collaborative learning interactions lasting ~ 60 min per session. The results showed that groups with different levels of shared understanding exhibit significant differences in time spent and maximum duration of referring and following behaviours. Further analysis using process mining techniques revealed that groups with different outcomes exhibit different *patterns* of group interactions. A loop between students' referring and following behaviours and resource management behaviours was identified in groups with better collaborative learning outcomes. The study indicates that the nonverbal behaviours studied here, which can be auto-detected with advanced computer vision techniques and multimodal data, have the potential to distinguish groups with different collaborative learning outcomes. Insights generated can also support the practice of collaborative learning for learners and educators. Further research should explore the cross-context validity of the proposed distinctions and explore the approach's potential to be developed as a real-world, real-time support system for collaborative learning.

**Keywords** Learning Analytics · Collaborative Learning · Process Mining

Extended author information available on the last page of the article

Published online: 28 November 2023

## 1 Introduction

Collaboration is a philosophy of interaction where individuals take charge of their actions (Laal & Ghodsi, 2012). It encourages sharing of authority and acceptance of the actions among group members, as well as consensus building through working together. Collaborative learning (CL) asks participants to apply this philosophy to live, and deal with other people in different educational contexts, such as in the classroom, in group meetings, and within their families (Panitz, 1999). The benefits of collaborative learning are various, including learning how to resolve social problems (Johnson et al., 1985), developing social interaction skills (Cohen & Cohen, 1991), building more positive heterogeneous relationships (Webb, 1980), encouraging diversity understanding (Swing & Peterson, 1982), and helping students to resolve differences in a friendly manner. However, simply asking students to work together as a group on a task does not guarantee success in collaboration (Summers & Volet, 2010). Effective collaborative learning requires meaningful interactions among individuals in a group.

Multiple research approaches and methods are developed to investigate interactions among members in collaborative learning activities. As a research area focusing on the measurement, collection, analysis and reporting of data about students and the learning environments with the purpose of understanding and optimizing learning (Siemens & Baker, 2012), learning analytics is also used to investigate collaborative learning. More specifically, it has been used to investigate the different modes of collaboration (Reimann et al., 2011), predict collaboration performance (Gašević et al., 2015; Spikol et al., 2017), evaluate and measure collaboration quality (Khan, 2017), and provide adaptive support for groups to meet their aims in collaborative learning (Kumar et al., 2007). More recently, the wide use of different sensors also enables researchers to analyze collaborative learning via various modalities of data, such as audio signals (Lubold & Pon-Barry, 2014), video recording analytics (Cukurova et al., 2020), and biomarkers (Dikker et al., 2017). Furthermore, an increasing number of novel analytics methods have been applied to make sense of these data, such as social network analysis (Gašević et al., 2019), epistemological network analysis (Sullivan et al., 2018), and process mining techniques (Fan et al., 2021; Zhou et al., 2022). With further developments in access to data, as well as methods to process these data, learning analytics provide more possibilities for understanding, interpreting, and supporting collaborative learning in different contexts.

However, using learning analytics to explore collaborative learning is also frequently critiqued that their potential is too limited to interpret and model meaningful interactions required to achieve complex phenomena such as building shared understanding and taking actions to solve problems together. Dillenbourg, (1999) stated that effective collaborative interactions should be interactive, synchronous, and negotiable. Yet, it is hard to detect such interactions in collaborative learning only focusing on the outcomes of collaboration or looking only at logs of interaction data from online learning environments. Most of the existing analytics studies mainly investigate the outcomes or performance of collaborative

learning with computationally observable features extracted from raw data but neglect the importance of educationally meaningful interactions among learners. This research may lead to models for predicting the outcomes of collaborative learning with high accuracy, yet it does not provide insights into how exactly collaborative learning skills can be improved (i.e. Spikol et al., 2018). Alternatively, it might lead to research studies that present models of collaboration from various analytics, but these analytics' completeness and representation of the complex process of collaboration may easily be challenged. In learning analytics and AI in Education research, there is a need to move away from investigations of collaboration that merely focus on the outcomes and data representations that are far off proxies of what educational researchers and practitioners are interested in to support meaningful collaborative learning interactions.

Indeed, many studies from research communities like CSCL try to identify such interactions in the process of collaborative learning (Vuopala et al., 2016). Nevertheless, they tend to rely heavily on the analysis of data with manual coding. In addition, the use of cumulative measurements, which only calculate the mean of specific features in the whole process of collaborative activities, is the main trend of work. These accumulated measures can hardly represent the complex *process* of collaboration. Groups with similar cumulative measurements may present different patterns of group interactions during their collaboration which would have different implications on what feedback they should receive and how their collaboration practice can be further improved. Thus, there is a need to explore collaborative learning as a process (Kent & Cukurova, 2020). These methods consume a large amount of time, require significant research expertise, and are hard to be implemented real-time in teaching and learning settings. Therefore, the identification of meaningful interactions through multimodal data collected from real-world collaboration settings with advanced AI approaches might provide unique opportunities to address some of these issues.

With these research gaps in mind, this research conducts a study in a real-world face-to-face collaborative learning setting to identify educationally meaningful non-verbal group interactions using video and audio data modalities. Machine observable behaviours are first identified using engineered features from computer vision and speaker diarization. Next, statistical comparison tests and process mining techniques are applied to explore how different groups' interaction patterns differ and emerge during collaborative learning. At last, the observed patterns are discussed with relevant collaborative learning theories and considerations with regard to their contribution to the theory and practice of collaborative learning.

## 2 Background

In the past decade, an increasing number of researchers focused on using learning analytics (LA) to detect meaningful interactions during collaborative learning. LA provides opportunities to collect data about learners and learning contexts to understand the process of collaborative learning, gain a deeper interpretation of collaborative learning, predict students' performance (Pérez Sánchez et al., 2022), and provide

support accordingly (Lias & Elias, 2011). With the development of information technology, different types of data have been collected from collaborative activities to generate a deeper understanding of collaborative learning, such as video data, audio data, log data, and physiological data. Using a variety of modalities of data, Multimodal Learning analytics (MMLA) extends LA by integrating and triangulating these data from different sources to quantify, infer and model complex learning processes (Worsley & Blikstein, 2011). For example, Oviatt et al. (2018) used digital pens to collect data about students' writing behaviours and combined them with video and audio data to distinguish the different performance exhibited by domain expert learners and non-expert learners in collaborative learning activities. Similarly, Spikol et al. (2018) collected video data, audio data and digital traces from collaborative problem-solving activities and used them to predict the quality of collaboration. This line of research tends to apply "black-box" methods which directly connect derived features extracted from raw data with the learning outcomes or quality of collaborative learning (initially judged by human experts/practitioners as the ground truth) to be predicted, but often overlooks meaningful interactions that occurred during the process which might be used to generate insights for feedback.

Recently, in order to open these "black box" approaches, a multi-layer framework which indicates an educationally meaningful method of mapping raw data to high-order constructs has been proposed for collaborative learning analytics (Martinez-Maldonado et al., 2021; Wise et al., 2021). There are five layers that consist of the framework, namely data, derived features, behavioural markers, sub-constructs and higher-order constructs. Derived features refer to the computationally detectable measurements extracted from raw data. Behaviour markers are the human observable behaviours which are related to the learning construct studied. Sub-constructs are the indicators which cannot be directly observed but are related to the educationally meaningful constructs for collaborative learning. It is argued that this framework can help to conceive and implement connections between data and learning constructs (Wise et al., 2021). Also, it may help with overcoming the problem of transparency and interpretability which appears particularly prominent in modern machine learning and deep neural network methods (Cukurova et al., 2020).

### 2.1 Nonverbal interactions of collaborative learning

In collaboration analytics, these derived features and behaviour markers tend to rely heavily on investigating students' verbal interactions. This may be due to the better interpretability of complex mechanisms of collaboration with verbal data, the particular focus of researchers on the cognitive dimension of collaboration and the potential accessibility of it through students' thoughts from their verbal outputs. However, the reliability and validity of verbal interactions on their own to holistically interpret the complex socio-cognitive and affective phenomenon of collaborative learning may be limited (Cukurova et al., 2018). Meanwhile, the analysis of verbal data usually requires manual transcript and coding, which is hard to avoid the subjectivity of the coder. Even though the applications of natural language processing (NLP) techniques enabled automatic analysis of verbal data, it still relies on

non-semantic probability calculations or qualitative value judgments to give meaning to each vocabulary vector based on the labels provided by humans. Although the contribution of verbal interactions' analysis to collaboration analytics cannot be denied, investigations of non-verbal aspects of collaboration are undervalued and understudied. During collaboration activities, nonverbal behaviours convey information not only to each individual who participates in collaboration but also reveal the nature and quality of interactions in collaboration through the analysis of behavioural cues extracted from non-verbal data to produce social signals. Vinciarelli et al. (2009) summarized five types of behavioural cues frequently used for social signal processing, namely physical appearance, gestures and posture, face and eye behaviour, vocal behaviour, and environment. These cues can be used to understand and interpret human social states such as emotion, attitude, physical interactions, and emblems. Among the presented five types of behavioural cues, vocal behaviour and eye behaviour can indeed be used to support collaboration analytics. For instance, Zhou et al. (2021) illustrated that detecting and visualizing students' speaking time and turn-talking behaviours in online synchronous meetings can support them to be aware of each member's participation during the collaboration and take actions to promote equal contributions to group discussions accordingly. Also, there are many studies which revealed the close relationship between gaze behaviours and the quality of collaboration in both digital collaborative learning environments (Schneider & Pea, 2013) and face-to-face collaborative learning activities (Schneider et al., 2021).

### 2.2 Process analyses of collaborative learning interactions

After the generation of derived features and behavioural markers, these can be analysed with multiple methods to reveal differences in collaborative learning outcomes. Although, most studies in the literature use statistical analysis of cumulated values of derived features and markers, only using cumulative measurements to analyze the process of collaborative learning is limited. Traditional research methods in these approaches, such as correlational analysis, regression, hierarchical linear modelling, can help us model and monitor the nonlinear and dynamic elements of collaborative learning to a certain extent (Amon et al., 2019; Vogler et al., 2017). Collaborative learning is a dynamic, multimodal, and synergistic process, in which interactions, cognitive development, and regulation influence each other dynamically (Stahl & Hakkarainen, 2021; Vogler et al., 2017). However, there is a lack of emphasis on researching the dynamic and temporal elements of collaborative learning, which may cause an oversimplified representation of the complex process of collaborative learning (Ouyang et al., 2022). For example, joint visual attention (JVA) is a concept which describes the average frequency of a group of students gazing at the same area in their working space. It is shown to have close correlations with the quality of collaboration (D'angelo & Schneider, 2021; Sharma et al., 2021). However, cumulative measurements, which usually present the average frequencies of specific behaviours or means of the detected metrics during the whole collaboration process, cannot present any insight into how individual learners use their gaze behaviours during the collaborative activity, in what sequence and through what kind of process. The sequence

of interactions plays a significant role in the success of collaboration, even though the accumulated results of these sequences might not present any statistically significant differences (Zhou et al., 2022). In addition, temporal considerations of when exactly a particular sequence of actions happens are often overlooked in cumulative evaluations. Without such considerations, teachers and researchers can have an overview of students' performance, but might still not know when and how exactly they should provide support in collaborative learning. Previous research showed that teachers might use the details of the collaborative learning process to have a deeper level of evaluation beyond the overall quality of collaboration (Alzahrani et al., 2023).

Recently, the wide use of process mining techniques provides new opportunities to gain more details about the process of collaborative learning. In the context of individual learning, some researchers applied Hidden Markov Models or First-order Markov Models to reveal the process of students' self-regulated learning (Fan et al., 2021). These techniques help distinguish students with different levels of learning performance in terms of their learning process. In collaborative learning contexts, sequence analysis and process mining were argued to provide opportunities to understand temporality in the learning process (Chen et al., 2022). Yet, there are fewer studies that tried to apply process mining techniques to analyze the process of collaborative learning. In their previous work, Schoor and Bannert, (2012) tried to explore the process patterns of working on the task, monitoring, and coordinating in collaborative learning by using process mining techniques. They have identified a double loop of these socially-shared regulations and found this in both high- and low-achieving dyads. Similarly, Zheng et al. (2023) used the lag sequential analysis method to compare the co-regulated behavioural patterns between students with and without learning analytics feedback in collaborative learning. It helped to illustrate that learning analytics feedback can foster students' process of co-regulation. However, both studies were conducted in a digital collaborative learning context.

Our goal in this paper is to extend the above literature by identifying meaningful interactions during collaborative learning through the analysis of nonverbal gaze and speaking behaviours extracted from video and audio modalities that are machine observable. We aim to answer two main research questions posed below:

1) What meaningful group interactions can be detected from machine observable nonverbal gaze and speech behaviours from video and audio data of students' face-to-face collaborative learning?
2) What are the variations in detected gaze and speech behaviours of groups with different collaborative learning outcomes?

## 3 Context

### 3.1 Educational context

The study was conducted in a postgraduate module at a tertiary level institute in the UK. Thirty-four students were divided into groups of four or five. The groups were assigned by the teaching team with considerations to create mixed-gendered,

interdisciplinary, and varied first language groups. Before the study started, ethical approval was received from the institution and individual consent was given by students after they were provided with a detailed information sheet about the project's goals, specific data modalities collected and how these data modalities are processed. Participation had no relation to the summative assessment of the module. Students were able to opt out at any time during the study.

During the 10-week course, students were asked to collaborate in groups to design a technological solution to overcome an educational challenge chosen by the group. For each week, students had to participate in face-to-face group discussions to work on their weekly collaborative group tasks. During the face-to-face sessions, they were asked to finish the tasks on Miro (a collaborative, digital design thinking platform), which were designed to scaffold their design work. Students were also able to ask help from teachers if they have any questions about the Miro tasks. The face-to-face sessions usually lasted for 60 min, but students could leave early if they finished their tasks or stayed a bit longer if needed. At the end of the term, students were asked to give a presentation about their group design work but no summative assessment was based on the collaborative activities apart from the formative feedback they received on their content and presentation. The summative evaluations of the module were writing tasks, both of which, benefited from inputs from students' collaborative learning tasks.

### 3.2 Data collection

During the face-to-face group discussions, students were seated as a group around a T-shaped table (see Fig. 1. below). In total, there were seven groups sat in one classroom. Each student accessed an online collaboration platform, Miro, through their own laptop/tablet. An Intel RealSense camera was set at the far end of the T-shaped table to capture video data. The video data was captured as.bag files with the frame rate of 30 fps and then has been transferred to.mp4 files. Meanwhile, a Boya conference microphone was used to record the group discussions during the session. OBS Studio, a streaming app, was used to help with the noise cancellation for the audio data while data collection. The audio data was

**Fig. 1** Data collection setting

collected as.mp3 files. Since the video data and audio data were not collected synchronously, timestamps from the meta-data were used to synchronize the videos and audio. Due to technical and ethical issues, a total of twenty collaborative learning sessions, lasting from about 33 min to about 67 min, have been analysed in this study.

### 3.3 Evaluation of collaborative learning outcomes

Self-report data was used to evaluate the students' insight towards the process of collaboration. After each group discussion session, students were asked to fill in a post-survey with 5-point Likert scale questions about their shared understanding using items from Kormanski (1990) and satisfaction with collaborative learning inspired by Dewiyanti et al. (2007). Based on the average scores of all students in the same group, the sessions were divided into groups with high and low shared understanding (SU) and high and low satisfaction towards collaboration (SC) experience groups. Meanwhile, two educators evaluated the weekly group outputs based on their knowledge mastery, task completion, and the complexity of the products. Similarly, the analysed sessions were divided into groups with high and low quality products (PQ) produced as collaboration outcomes based on the overall scores groups achieved across these three dimensions.

## 4 Methodology

This study applied the framework of 'from clicks to constructs' to map low-level data collected from real-world learning settings to model and interpret meaningful collaborative learning constructs (Wise et al., 2021). Figure 2 shows an overview of the framework adopted and applied in this study which will be discussed below.

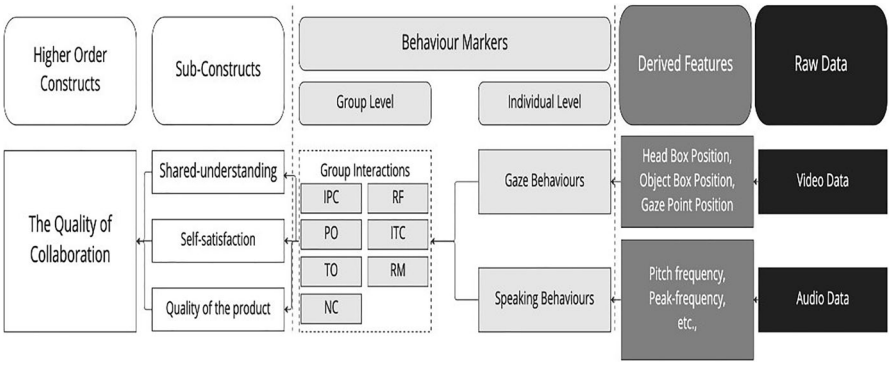


**Fig. 2** Mapping from the captured data (right) to the targeted constructs (left)

### 4.1 Extracting behavioural markers from multimodal data

#### 4.1.1 Speaker Diarization from the audio data

The audio data was uploaded to the Amazon Web Services (AWS) cloud and then automatically analysed by Amazon Transcribe which is an automatic speech recognition service. The output results were given as.json files, which contained the content, speaker ID, and time stamps of the start and end time of each speech detected in the audio data. Then, the.json files were converted into.csv files which present the speaker of each second during the collaborative learning session. It is worth noting that, since there were no pre-recorded voice samples from students, initially manual work was needed to map the voice in the audio data to the individual students.

#### 4.1.2 Gaze behaviour annotations from the video data

Four types of gaze behaviours, namely gazing at peers, gazing at laptops, gazing at tutors, and gazing at other objects, were identified from the video data. There are two main reasons for the decision of using these particular gaze behaviours. First, from a learning sciences point of view, these four gaze behaviours were illustrated to have the potential of distinguishing the process of collaborative learning for groups with different learning outcomes (Zhou et al., 2022). Second, from a machine learning point of view, they can be automatically detected with existing computer vision techniques with a high level of accuracy (Zhou et al., 2023). The first frame of each second from a particular session was extracted to generate a new video for labelling gaze behaviours. Computer Vision Annotation Tool (CVAT) (cvat.org) was used for video annotation. The coding work was conducted by two researchers. Before they worked separately, a sample video of 1000 frames were coded by both researchers to test the reliability of the process and reach a consensus on the coding process. Inter-rater reliability of double coding presented very high-reliability values (Cohen's Kappa = 0.98) and the rest of the coding was completed by two researchers individually.

### 4.2 Defining group interactions through gaze and speaking

This paper focused on six types of group interaction status derived from literature about collaborative learning. A rule-based method was used to identify these group interactions through the speaking and gazing behaviours detected from the multimodal data. This section will briefly define these group interactions and introduce how they were identified from the behavioural markers.

#### 4.2.1 Interaction with peers through communication (IPC)

Verbal communication between peers has been considered an important type of interaction since it can help students with building shared understanding and socially-shared regulation of collaborative learning (Ouyang & Xu, 2022). In this

study, interaction with peers through communication (IPC) is defined as the situations in which students were trying to build shared understanding through negotiation and discussion. It involves both oral expression and active listening. Therefore, if there were more than one student speaking and over half of the group members gazing at the speaker, the group status was coded as IPC.

#### 4.2.2 Referring and following (RF)

Another type of interaction defined in this study focuses on the discussion based on specific learning materials or learning activities. Gaze following has been considered as an act of moving one's gaze attention to the object which was gazed at or introduced by another person. From the perspective of neuroscience, gaze following is considered to be an important behaviour closely related to collaborative shared attention (Emery, 2000). In this study, referring and following (RF) is developed from the concept of gaze following. It is defined as situations when students in a group paying attention to their laptops (learning resources) based on one member's oral expressions. Thus, the RF code was used if one student was speaking while more than half of the group members were gazing at their laptops.

#### 4.2.3 Peer observation (PO)

Besides communication, observation is argued to be an important type of interaction during collaborative learning. It can reflect students' regulation dimension of monitoring (Ouyang et al., 2022) and is associated with the success of collaborative learning (Cukurova et al., 2020). In this study, peer observation (PO) refers to the situations in which students try to understand other student(s) from the behavioural dimension by looking at this student. The behavioural dimension refers to one student trying to understand what actions were taken by others. It may occur when students try to understand and make sense of one student's demonstrations. Therefore, if there were no students speaking and over half of the group members were gazing at peers, the group status was coded as PO.

#### 4.2.4 Interactions with a tutor through communication (ITC)

The interaction between students and tutors is also important in collaborative learning. It can support students with the understanding of domain knowledge as well as monitoring the collaborative learning process (Le et al., 2018). Interaction with a tutor through communication (ITC) was defined as a situation in which students were discussing with tutor(s). Code ITC would be used if there was a student speaking and over half of the students were looking at the tutor.

#### 4.2.5 Tutor observation (TO)

Tutor observation (TO) is another type of interaction between students and tutors and it refers to situations when students were actively listening to a tutor who might be explaining specific content, answering student questions or clarifying the

learning activities. Tutor observation (TO) is distinguished from interaction with a tutor through communication (ITC) because tutors may act as different roles in these interactions. In interaction with a tutor through communication (ITC), tutors take more responsibility for scaffolding the group work since they are actively exchanging their points of view with students through discussions with inputs from students. Yet, during TO, tutors are more likely to be a "presenter"/"lecturer" of the domain knowledge in a more traditional pedagogical approach (Le et al., 2018).

#### 4.2.6 Resource management (RM)

The interactions with other learning resources apart from peers and tutors (i.e. laptops, books, tablets etc.) also play a significant role in the effectiveness of collaborative learning (Ouyang et al., 2022). For instance, previous research in CSCL illustrated that groups with higher attention synchrony in virtual collaboration environments may achieve higher collaborative learning outcomes (D'angelo & Schneider, 2021). Furthermore, it was found that informing students about where other members were working in the virtual learning environments can promote the outcomes of collaboration (Schneider & Pea, 2013). This study considered resource management (RM) behaviours as situations in which students focused on the same learning materials or activities on their laptops/tablets. If no student was speaking while more than half of the group members were gazing at their laptops, the RM code was applied.

#### 4.2.7 Non-collaboration (NC)

Besides the six codes defined above, code "non-collaboration" (NC) was used when no group interaction status was detected in the analysed window. Based on the qualitative observation of original video data, the NC codes were mainly those times when students worked individually rather than having any interactions with other members.

Figure 3 shows an overview of the rules for the detection of group interaction status. All seven group interaction statuses were determined by individual students' gaze and speaking behaviours in a window of five seconds. After the detection of group interaction status, adjacent windows with the same encoding were merged as a new event. The start time and end time of the events were calculated based on the original windows.

Four measurements were used to calculate the group interaction status detected above. Freq_X was used to describe the average time a group status code X occurs in a minute. It presented how frequently group status X was applied during the collaborative learning process. TS_X was calculated as the average time that status X lasted in a minute. It showed how long one group's status X lasted in a minute. Mean_X was used to describe the average duration status X occurred in a group, while Max_X presented how long the longest status X lasted. These four measurements were applied to all seven group interaction statuses defined above. Therefore, twenty-eight variables were used to compare the groups with different levels of shared understanding (SU), self-satisfaction with the collaboration (SC) process, and the outcome they produced as the product quality of collaboration (PQ).

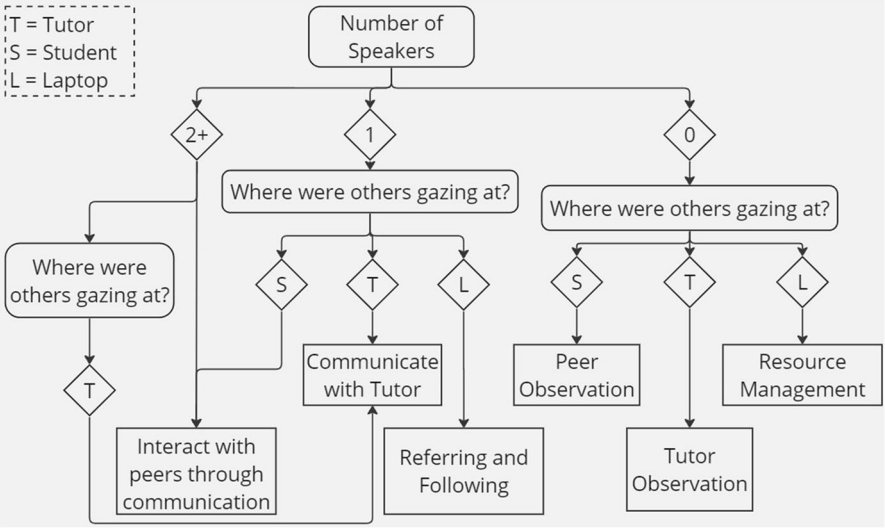


**Fig. 3** The rules applied for inferring group interaction statuses from gaze and speaking behaviours where S, T and L denotes student(S), tutor(s) and laptop(s), respectively

### 4.3 Statistical analysis

Before the process mining, conventional statistical analyses were conducted to explore RQ1 and RQ2. Firstly, descriptive analytics was used to show the distribution of time spent in different types of group interactions for each session. It investigates the extent to which the identified group interactions can be observed in collaboration sessions. Then, comparative analyses tests were used to explore the differences in group interaction status between groups with different Shared Understanding (SU), Satisfaction with Collaboration (SC), and Product Quality (PQ) outcomes. Normality tests and Homogeneity of variance tests were conducted to check the parametric assumptions. Based on the results of Normality tests and Homogeneity of variance tests, as well as the small scale of the sample size, Mann–Whitney U-tests were used to compare the differences between groups.

### 4.4 Process mining

Fluxicon Disco (https://fluxicon.com/disco/), a process mining tool, was used to explore the different processes of collaboration exhibited by groups with different levels of SU, SC, and PQ. It can show how one group moves between different group interaction statuses. Data from twenty sessions were merged together in an excel document. Each row of the document presented an event of group interaction status per five-second window. In addition to the status code as well as the start and end time of the status, each row also contained the session ID and the SU, SC, and PQ levels of the session. In total, 4126 rows of events were analysed in Disco for

process mining. The output of the process mining as flowcharts which show how students move between different group interactions were created.

## 5 Results

### 5.1 Statistical analysis

#### 5.1.1 Time spent in different types of group interaction status

Table 1 below shows the distribution of time spent in different types of group interactions for each group. The cell with 0% means that this particular group interaction status was not detected in the group. As Table 1 shows, four types of group interactions, namely interaction with peers through communication (IPC), peer observation (PO), referring and following (RF), and resource management (RM) were detected in almost all sessions. To be more specific, interaction with peers through communication (IPC) and referring and following (RF) are two types of group interactions in which most groups spent most of their time in. They can also be seen in all twenty sessions. Furthermore, although peer observation (PO) and resource management (RM) took less time in group interactions, they were also present in all groups but one. Moreover, interaction with a tutor

**Table 1** The distribution of time spent in different types of group interactions

| Session | IPC | PO | RF | ITC | TO | RM | NC |
|---|---|---|---|---|---|---|---|
| 1 | 39.17% | 32.10% | 10.75% | 0.92% | 0.15% | 14.75% | 2.15% |
| 2 | 70.46% | 1.36% | 16.80% | 9.76% | 0.27% | 0% | 1.36% |
| 3 | 28.18% | 10.57% | 33.33% | 0.27% | 0% | 24.12% | 3.52% |
| 4 | 58.25% | 0.17% | 35.69% | 3.03% | 0% | 0.17% | 2.69% |
| 5 | 59.74% | 1.30% | 35.71% | 0.32% | 0% | 1.79% | 1.14% |
| 6 | 50.00% | 2.37% | 33.39% | 4.75% | 1.27% | 7.91% | 0.32% |
| 7 | 32.43% | 0.49% | 53.56% | 0% | 0% | 9.09% | 4.42% |
| 8 | 47.80% | 1.13% | 38.64% | 1.63% | 0.13% | 7.28% | 3.39% |
| 9 | 22.73% | 7.24% | 33.52% | 0.71% | 0% | 29.97% | 5.82% |
| 10 | 65.36% | 0.16% | 27.61% | 0% | 0% | 3.10% | 3.76% |
| 11 | 41.07% | 0.35% | 45.06% | 0% | 0% | 8.84% | 4.68% |
| 12 | 43.63% | 0.29% | 43.20% | 3.00% | 0% | 7.73% | 2.15% |
| 13 | 53.58% | 4.15% | 34.72% | 0% | 0% | 3.40% | 4.15% |
| 14 | 62.26% | 0.38% | 28.30% | 0% | 0% | 1.51% | 7.55% |
| 15 | 49.78% | 0.44% | 33.14% | 0% | 0% | 1.91% | 14.73% |
| 16 | 56.00% | 0% | 33.74% | 0% | 0% | 6.43% | 3.83% |
| 17 | 48.51% | 3.53% | 39.59% | 0% | 0% | 6.13% | 2.23% |
| 18 | 51.24% | 2.23% | 17.87% | 15.51% | 2.11% | 1.24% | 9.80% |
| 19 | 38.64% | 2.49% | 37.53% | 0.14% | 0% | 9.00% | 12.19% |
| 20 | 40.97% | 1.48% | 39.76% | 0% | 0% | 14.02% | 3.77% |

through communication (ITC) was detected in eleven sessions while tutor observation (TO) was detected in only five groups. It may be because not all groups requested help from tutors during their collaborations. Lastly, it is also interesting that non-collaboration (NC) status was detected in all the twenty groups. This might indicate moments when the group members worked individually are also an important part of collaborative learning in real-world settings. This might also indicate that the interaction types we proposed and detected here can be further extended to capture other relevant group statuses as well.

#### 5.1.2 Statistical comparison between different SU groups

As the datasets failed parametric assumption tests, Mann–Whitney U-tests were used to compare the differences in group interaction status between groups with different Shared Understanding (SU), Satisfaction with Collaboration (SC), and Product Quality (PQ) outcomes.

Mann–Whitney U-test (Table 2) demonstrated that there was significantly higher time spent in referral and follow interactions (TS_RF) in the high SU group (Md = 23.22, $n = 10$) compared to the low SU group (Md = 20.04, $n = 10$), U = 20.00, Z = -2.27, $p = 0.02$, r = 0.51. It means that the groups with high SU levels spent more time in referring and following (RF) interactions than the groups with low SU levels. Also, the high SU groups (Md = 13.37) exhibited significantly higher Mean_RF than the low SU groups (Md = 10.82), U = 19.00, Z = -2.34, $p = 0.02$, r = 0.52. It means that the referring and following (RF) interactions detected from the high SU groups usually lasted longer than the low SU groups. In terms of Max_RF, the high SU groups (Md = 60.00) were also significantly higher than the low SU groups (Md = 40.00), U = 18.50, Z = -2.42, $p = 0.02$, r = 0.54. It illustrates that the longest referring and following (RF) events detected from the high SU groups were usually longer than the low SU groups. It is worth noting that,

**Table 2** Mann–Whitney U-test of TS_RF, Mean_RF, and Max_RF between the high and low SU group

| | High SU Groups | | Low SU Groups | | U | Z | *p* | r |
|---|---|---|---|---|---|---|---|---|
| | n | Median | n | Median | | | | |
| TS_RF | 10 | 23.22 | 10 | 20.04 | 20.00 | -2.27 | 0.02** | 0.51 |
| Mean_RF | 10 | 13.37 | 10 | 10.82 | 19.00 | -2.34 | 0.02** | 0.52 |
| Max_RF | 10 | 60.00 | 10 | 40.00 | 18.50 | -2.42 | 0.02** | 0.54 |

**. *p* is significant at the 0.05 level (2-tailed)

TS_RF: the average time that referring and following (RF) status lasted in a minute

Mean_RF: the average duration that referring and following (RF) status occurred in a group

Max_RF: the duration of the longest referring and following (RF) status

although groups with different SU levels emerged with significantly different time

spent, average duration, and extremum of referring and following (RF) status, the frequency of referring and following (RF) status occurred in the process of collaboration did not show any significant differences.

### 5.1.3 Statistical comparison between different SC groups.

Two significant results were observed from the Mann–Whitney U-test of groups with different levels of satisfaction from their collaboration experience (SC) (Table 3). For instance, groups with high SC levels (Md = 2.50, $n = 11$) had significantly higher TS_NC than groups with low SC level (Md = 2.04, $n = 9$), U = 22.00, Z = -2.09, $p = 0.04$, r = 0.47. It shows that high SC groups spent more time in individual work and resting status during the process of collaborative learning. Also, high SC groups (Md = 20.00) exhibited significantly higher Max_NC than low SC groups (Md = 15.00), U = 20.50, Z = -2.24, $p = 0.03$, r = 0.50. It shows the longest NC events, which presented the individual work time, detected from the high SC groups lasted longer than that detected from the low SC groups.

### 5.1.4 Statistical comparison between different PQ groups

In terms of comparison between groups with high PQ levels and groups with low PQ levels, there were no significant differences in any of the group interaction statuses detected. It shows that groups with different PQ levels seemed not to exhibit significantly different group interaction statuses in accumulated measures we used in statistical analysis. Statistical analyses which focus on cumulative measurements are by nature limited to be used to interpret the interaction differences in the complex process of collaborative learning. Therefore, the next section will present how groups with different SU, SC, and PQ levels change between different group interaction statuses during the process of collaboration.

**Table 3** Mann–Whitney U-test of TS_NC and Max_NC between the high and low SC group

| | High SC Groups | | Low SC Groups | | U | Z | *p* | r |
|---|---|---|---|---|---|---|---|---|
| | n | Median | n | Median | | | | |
| TS_NC | 11 | 2.50 | 9 | 2.04 | 22.00 | -2.09 | 0.04** | 0.47 |
| Max_NC | 11 | 20.00 | 9 | 15.00 | 20.50 | -2.24 | 0.03** | 0.50 |

**. *p* is significant at the 0.05 level (2-tailed)

TS_NC: the average time that non-collaboration (NC) status lasted in a minute

Max_NC: the duration of the longest non-collaboration (NC) status

## 5.2 Process mining

### 5.2.1 Comparison of the process of collaboration between different SU groups

Figure 4 presents the process maps for both high and low SU groups. As Fig. 4a shows, there were strong connections between referring and following (RF), peer observation (PO), and resource management (RM) observed from groups with high SU. These three types of group interactions shaped a loop in which students started with discussions based on the learning materials on laptops, sometimes followed by peer observation (PO) and then moved to take actions on their laptops to take individual actions based on the discussions and observations. The process map also shows a strong link from resource management (RM) to interaction with a tutor through communication (ITC). It illustrates that groups with high SU levels usually asked for help from a tutor after taking specific actions for a certain amount of time. Furthermore, it is interesting to observe from the high SU groups that peer interactions through communication (IPC) usually alternate with "individual work" (NC), which is followed by interactions with tutors (ITC).

In terms of the low SU groups, as Fig. 4b shows, there are three types of group interactions followed by discussions with tutors (ITC), namely referring and following (RF), resource management (RM), and tutor observation (TO). It also can be observed that low SU groups frequently moved from tutor observation (TO) to discussion with peers (IPC). These illustrate that the interactions with a tutor in low SU groups usually triggered group discussion or taking actions to work on the learning activities. Yet, different types of interactions between students did not show strong connections with each other. This may illustrate that low SU groups needed tutors' support to help them build a common understanding towards the learning contents and learning activities.

Compared to the low SU groups, high SU groups tended to ask tutors for help after taking actions on the learning activities rather than communicating with tutors first and then taking actions accordingly. Furthermore, high SU groups exhibited stronger connections between different types of group interactions among peers.

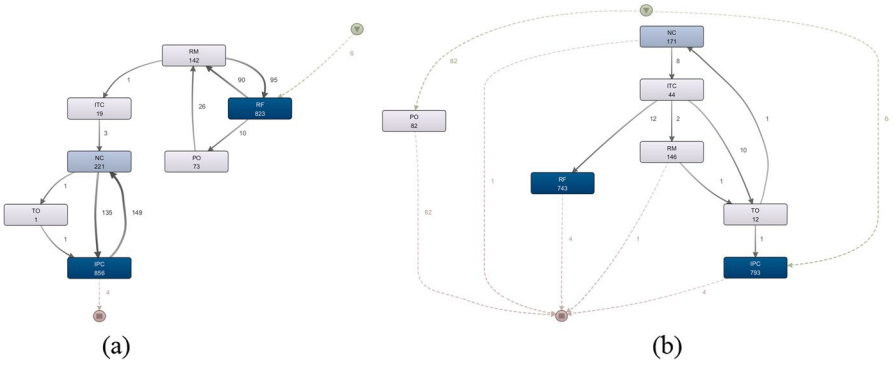


**Fig. 4** Process maps of (**a**) high SU groups vs (**b**) low SU groups

They tended to use the loop of taking actions, observation, explanation and discussion as a strategy to co-regulate their group work. In contrast, this kind of strategy was hardly observed in low SU groups. They tended to focus on a single type of interaction among peers until the sessions ended. However, both groups with high SU levels and low SU levels also exhibited a commonality. For instance, interaction with tutors through communication (ITC) usually appeared with interaction with peers through communication (IPC) in the same paths (e.g. ITC → NC → IPC, and ITC → TO → IPC). This endorses that teachers' support in these sessions is likely to foster students' discussion during collaborative learning.

### 5.2.2 Comparison of the process of collaboration between different SC groups

Figure 5 shows the process maps of high and low SC groups. According to Fig. 5a, groups with high SC levels exhibited strong links between referring and following (RF) and resource management (RM) statuses. It means that students from high SC groups frequently switched between these two types of peer interactions. They started with a discussion on the specific materials on their laptops and took actions based on the discussion. Then, a new round of discussions was carried out to reflect on the work they have done. This cycle appeared many times until students were satisfied with their work.

As Fig. 5b shows, groups with low SC levels had clear connections between interaction with a tutor through communication (ITC) and tutor observation (TO). It illustrated that students from low SC groups frequently raised a question to tutors or confirmed their thoughts with tutors through mutual communication. Then they spent a certain amount of time listening to tutors for detailed explanations. Until

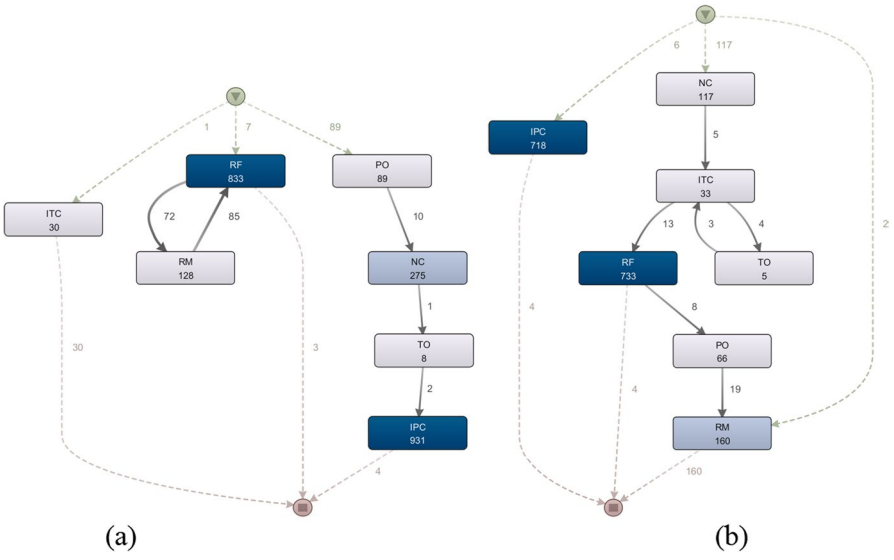


**Fig. 5** Process maps of (**a**) high SC groups vs (**b**) low SC groups

they finished interactions with tutors, they moved to referring and following (RF) status for further discussion based on the materials on their laptops and then took actions for the group work accordingly.

Although both groups with high and low SC levels exhibited connections between different types of interactions among peers, high SC groups seemed to use a strategy of iteration in which transitions happened many times between referring and following (RF) and resource management (RM) status. In this iteration of discussion and action taking, students appear to become more satisfied with their group work. In contrast, low SC groups tended not to take actions unless they achieved consensus through communications with tutors and peers. Meanwhile, they did not go back and discuss the work after they took actions. This might be the reason why they have a lower SC level. Since there was little to no discussion on the finished work, they could hardly reflect on the work and adjust it toward a higher satisfaction level of their final products.

### 5.2.3 Comparison of the process of collaboration between different PQ groups

The process maps of high and low PQ groups are shown in Fig. 6. Based on the process maps, high PQ groups tended to start with interaction with peers through communication (IPC), and then spent time on non-collaboration (NC) before moving to a cycle of referring and following (RF) and resource management (RM). To be more specific, the high PQ groups first had mutual communication among peers to discuss the learning activities. Then, before they moved to a more specific discussion, they tended to have some rest or individual work. After that, they looked at their laptops and had more focused discussions alternated with taking actions for group work.

In contrast, groups with low PQ levels usually started by taking actions for their group work. Then, once they met problems or challenges in their group work, they tended to

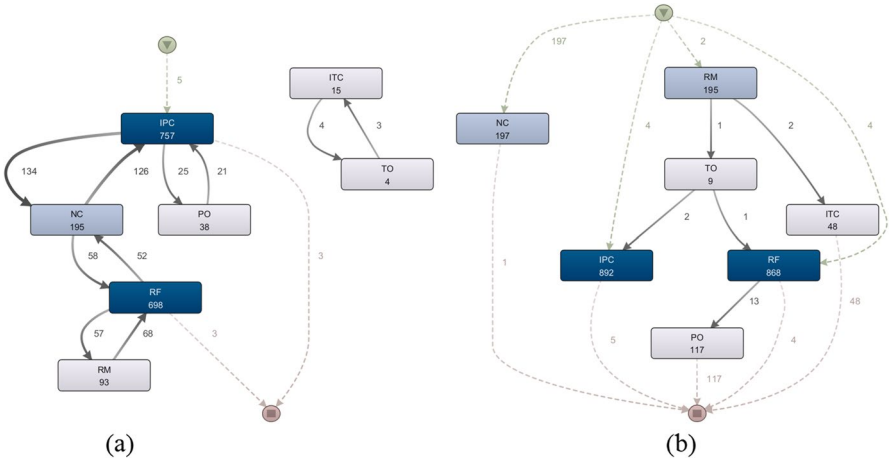


**Fig. 6** Process maps of (**a**) high PQ groups vs (**b**) low PQ groups

ask for help from tutors. Therefore, Fig. 6b, resource management (RM) status was followed by interaction with a tutor through communication (ITC) and tutor observation (TO). These interactions with tutors then led to more interactions among peers, such as interaction with peers through communication (IPC) and referring and following (RF).

By comparing the high and low PQ groups, it is found that they seemed to use different strategies during the process of collaborative learning. High PQ groups tended to discuss together towards a plan before they took actions. Also, while they were taking actions, they may have reflected on their work through discussions and made changes accordingly. On the other hand, the low PQ group tended to start with taking actions and then interacted with tutors and peers to discuss the difficulties or challenges arising from it. It is also worth noting that, similar to the high SU groups and high SC groups, the cycle of referring and following (RF) and resource management (RM) occurred frequently in the process map of high PQ groups. The iteration of taking actions and reflecting on these actions in a cyclical manner does not only appear to be related to higher levels of shared understanding and self-satisfaction, but it may also lead to better product outcomes through collaboration.

## 6 Discussion

Using multimodal learning analytics in physical collaborative learning has been studied for some years now. However, most research focused on the automation of collecting relevant multimodal data from the learning environments, predicting the learning outcomes, as well as modelling the learners and learning processes (Chua et al., 2019). How to apply existing research results to real-world collaborative learning practice is still under-explored which makes the real-world impact of these analytics solutions limited (Alwahaby et al., 2022). On the one hand, limited human interpretability of the commonly used data logs for predicting and modelling collaboration in existing works makes it hard to provide educationally and pedagogically valuable information. On the other hand, existing studies tend to distinguish the groups with different learning performances through the comparison of cumulative measurements, which leads to the differences in the order and pattern of learning behaviours during collaboration to be overseen. Both of these reasons make it a challenge for existing research to support teachers and students in a real-world collaborative learning context. Therefore, this study aims to identify pedagogically meaningful group interactions from machine observable nonverbal behaviours from audio and video data and uses process mining to investigate how groups differ in their interactions during the process of collaborative learning.

In order to answer RQ1: *What meaningful group interactions can be detected from machine observable nonverbal gaze and speech behaviours from video and audio data of students' face-to-face collaborative learning?* This study first extracted behavioural markers about gazing and speaking from video and audio data. Based on these behavioural markers, seven types of group interaction status, namely interaction with peers through communication (IPC), peer observation (PO), referring and following (RF), interaction with a tutor through communication (ITC), tutor observation (TO), resource management (RM), and non-collaboration (NC) have been identified. The results show that all three interactions between peers, interaction with peers through communication

(IPC), peer observation (PO), and referring and following (RF), can be observed in most analyzed sessions. To be more specific, interaction with peers through communication (IPC), which demonstrates the situations during which learners in a group have mutual discussions, has been observed in all sessions and occupied the biggest portion of time during collaborative learning. This aligns well with previous research which shows the importance of interactions through oral communication during collaborative learning (Cukurova et al., 2020; Spikol et al., 2018). Furthermore, referring and following (RF) was another type of group interaction which was observed frequently in all sessions. It may refer to situations in which one learner was presenting their thoughts to others based on the content on their laptops/tablets and others followed. In addition to interactions between peers, resource management (RM), which represents situations in which students focus on their laptops/tablets to take actions for collaborative learning activities, also was detected in most sessions. This can be related to research which stressed the importance of individual accountability (Slavin, 1991). Individual accountability refers to students in a group to undertake their share in completing the task as well as acknowledging and promoting others' share. In this study, individual accountability was observed as actions taken on the collaboration platform and paying attention to input from others at the same time. Resource management (RM) might be a representation of individual accountability in which most students focused on fulfilling the tasks in their shared learning space. Moreover, besides these types of interaction statuses, there are also two statuses which only were detected in some of the sessions, namely interaction with a tutor through communication (ITC) and tutor observation (TO). This may be because not all groups interacted with tutors during the collaborative learning activities. Therefore, they did not initiate any interactions with the tutors. Lastly, the results show that all groups exhibited non-collaboration (NC) status during the process of collaboration. It means that none of these sessions only maintained positive group interactions at all times and students spent a certain amount of time concentrating on individual activities during the collaborative learning task. It is worth noting that, the interactions analysed in this study can be automatically extracted from the collaborative learning process using computer vision approaches. To what extent these auto-extracted interactions can reflect the quality of collaboration processes is still an open question that requires further elaboration and research. In this paper, we attempted to make connections to learning sciences literature on effective collaboration processes, yet some of these connections require further investigations. Future research should also conduct triangulation of the insights from machine observable features of collaborative interactions with other methods used to judge collaboration quality (e.g. expert observations, think aloud processes, interviews etc.)

Regarding RQ2*: What are the variations in gaze and speech interaction behaviours of groups with different collaborative learning outcomes?* We used both statistical analyses and process mining techniques to explore the differences in groups with different learning outcomes in terms of shared understanding, satisfaction and product quality. The results of the comparative analyses show that groups that spent more time on referring and following (RF), and had a longer average duration of referring and following (RF) status, might have a higher level of self-reported shared understanding. Given the fact that referring and following (RF) is developed from the behaviours of gaze following, it represents the situations in which one student spoke and

other students paid attention to the same materials on their laptop according to the speech. It does not only show the synchrony of the gaze behaviours but also indicates the active listening behaviours. By analyzing the verbal interactions, previous research from CSCL considered the referent or repeated words used by the students as an important element to indicate shared understanding (Stahl, 2002). The use of such utterances relied heavily on mutual dialogue which requires active presenting and listening (Bohm & Weinberg, 2004). Therefore, this result illustrates that the presentation of personal understanding and others' active listening may lead to the establishment of shared understanding in collaborative learning. It is also worth noting that the frequency of the referring and following (RF) status that occurred in the process of collaboration did not show any significant differences. This indicates that groups with a higher level of shared understanding tended to apply referring and following (RF) as a method of conducting long discussions rather than showing referring and following (RF) behaviours more frequently. referring and following (RF) interactions require a certain amount of time for students to present their points of view. These findings align well with the previous research which emphasized the importance of interactions such as sharing knowledge, explaining understanding, challenging others' opinions, and providing feedback to others for effective collaborative learning (Laal & Ghodsi, 2012; Le et al., 2018). Referring and following (RF) interactions are likely to involve moments in which students engage in expressing their own opinion. Furthermore, the comparative analysis also found that the groups which spent more time on the non-collaboration (NC) status might be more satisfied with their group work. It means that groups in which students spent more time taking actions individually were more likely to be satisfied with their final products even though this did not necessarily lead to better group products or shared understanding. Based on the existing studies, taking individual actions helps students establish and take personal responsibility which is also considered a key element of successful collaboration (Cukurova et al., 2018). Moreover, there was no statistically significant relationship between group interactions and the quality of collaboration products. There might be various reasons for this result. First, this may be because the quality of products produced by each group might not be only related to the quality of collaboration interactions, but also related to multiple intraindividual factors which we didn't include in this study (i.e. students' domain knowledge). Second, this may illustrate the limitation of the approach of analysing only non-verbal behaviours in collaborative learning. Whether students' interactions can contribute to their collaboration, or not, still heavily depends on the content of their communication rather than the mere fact that they are communicating. Therefore, non-verbal behaviours sometimes may fail to indicate that the extent to which interaction contribute to the quality of collaboration. For instance, students may have off-topic discussions during collaboration but this is unlikely to be distinguished from more meaningful discussions observing only the non-verbal behaviours. Therefore, there is a need to combine non-verbal behaviours with content analysis of verbal interactions to have a more comprehensive understanding of students' interactions.

The order of group interactions also varied between groups, those with high shared understanding tended to take actions before seeking help from tutors. Similarly, groups with high product quality tended to discuss more before taking actions, while those with low product quality tended to take actions first and discussed

problems later. Process mining techniques provided a more detailed understanding of collaborative learning interactions compared to traditional statistical analysis.

The process mining results revealed significant differences and similarities between the groups, with a loop between monitoring progress (RF) and taking actions (RM) observed in groups with high levels of shared understanding, self-satisfaction, and product quality. Although it is a multidimensional and much more complex phenomenon, this loop may be interpreted as a crude proxy of socially shared-regulation of learning in non-verbal behaviour markers we detected. Socially shared-regulation of learning (SSRL) was defined as a process of enacting a joint goal, monitoring progress toward the goal and regulating the learning process through making adjustments to cognitive, emotional, motivational and behavioural states as needed (Hadwin & Oshige, 2011). In this study, referring and following (RF) and resource management (RM) behaviours might indicate some of the monitoring phase actions since during referring and following (RF) students discussed existing input on their Miro boards, while resource management (RM) was taking actions to make adjustments accordingly. By transferring between taking action and discussing, students might be continuously reflecting on their existing work together and improving it over time. This aligns with previous research which suggests that co-regulated learning or socially-shared regulation of learning might help groups to build better-shared understanding and achieve better learning outcomes (Panadero & Järvelä, 2015). Furthermore, the order of different types of group interactions also shows differences between groups. To be more specific, groups with a high level of shared understanding tended to take actions on the learning activities before they asked for help from tutors. In contrast, groups with a low level of shared understanding tended to ask for support from tutors before they move to discussion or take actions on the learning activities. This may be because these groups needed further support from tutors to achieve an initial shared understanding before they move on. Similarly, the groups that produced high-quality products as a group usually discussed more before they took actions, while groups with a low level of product quality tended to take actions first and then had a discussion when problems occurred. As previous studies showed, students can benefit greatly from achieving consensus before they implement their plan in collaborative learning activities (Panitz, 1999). Groups with high product quality tended to conduct discussions to come up with a plan which is agreed upon by the group members before they carried it out. These results also illustrate the value of using process mining techniques to provide insights into students' group interaction during the process of collaboration. Such information on "how" cannot be generated using traditional statistical analysis, but it is essential to understand the ways in which the collaboration process can be improved with feedback.

This study makes both theoretical and practical contributions to collaborative learning research. It proposes a method for analyzing collaborative learning through the detection of non-verbal interactions using the "from clicks to constructs" framework (Wise et al., 2021) implementing it for machine observable speech and gaze behaviours. We identified seven types of educationally meaningful group interactions from the speaking and gazing behaviours of students and applied both statistical and process mining approaches to explore how these interactions were used by different groups.

The results are interpreted with previous literature on collaborative learning theories to judge the extent to which they are aligned or contradicted. Specifically, the loop of resource management (RM) and referring and following (RF) has been illustrated to be closely related to shared-understanding, satisfaction towards the products of collaboration, and the expert-evaluated quality of collaboration products. Furthermore, sessions with different collaboration outcomes also showed differences in the sequence of group interactions. This result contributes to the broader research in collaboration analytics to consider using process mining to analyze and model the complex system of collaborative learning with meaningful interactions. For instance, previous studies which focused on exploring the different phases of SSRL relied heavily on the analysis of students' verbal behaviours. The patterns of non-verbal behaviours identified in this study may provide more possibilities for identifying SSRL phases from non-verbal behaviours. Moreover, the study provides insights into how groups with higher levels of product quality tend to choose specific sequences of group interactions, which may inform instructional design and feedback interventions for collaborative activities. For example, the information about how groups with different learning outcomes interacted with teachers may suggest the value of teachers providing support *after* students' own exploration rather than immediately intervening with guidance. Also, the results of this study can be used to help with the design of more detailed collaborative learning activities to scaffold effective learning strategies.

Moreover, previous research stressed the challenges of ethical and privacy considerations in the real-world use and adoption of multimodal LA and AI approaches in Education (Alwahaby, & Cukurova, 2023; Alzahrani et al., 2023). It is argued that both teachers and students may be worried about the unreliable predictive results generated by "AI" (Seo et al., 2021). In addition, students may be concerned about being monitored by AI systems and might have performance anxiety as well as surveillance fears. For instance, Seo et al. (2021) reported students' worries about being judged by what they said unconsciously. Similarly, Zhou et al. (2021) found that some students feel uncomfortable with the analytics of their discussions and raised concerns that they can be hesitant to speak not to make mistakes during collaboration observations with AI and analytics. Despite their potential benefits, given the pervasive nature of multimodal data and opaque AI techniques that may be employed to process them, this line of research and practice present some significant ethical concerns (Alwahaby, & Cukurova, 2023). Detailed discussion of such concerns is not within the scope of this work, yet ethical issues in the adoption and use of AI in educational settings require careful considerations. In this study, we attempted to detect explainable non-verbal behaviours that can help address some of the ethical concerns thanks to their transparent nature. However, a myriad of other potential ethical issues (e.g. fairness, agency, accountability, surveillance) that might influence the real-world use and adoption of computer vision approaches in educational settings require future research and elaborations.

**Limitations** Finally, it is important to note some limitations of this study. First, the causal relationship between group interactions and collaborative learning outcomes is unclear. Further research is needed to infer whether the groups achieved higher SU and SC levels because of the specific sequences of group interactions they have

applied, or that these group interactions are the results of groups with higher SU and SC. It is likely that these relationships are more dynamic rather than monodirectional cause and effect. During the process of collaborative learning, the outcomes and the way of interaction are likely to mutually influence and promote each other. It is worth exploring further the mechanisms of this process in more tightly controlled studies. Second, relying solely on self-reported questionnaires and evaluations of the final products of collaboration by researchers may limit the evaluation of collaborative learning outcomes studied. Future studies may consider using different evaluation frameworks from CSCL research to assess the collaboration process (i.e., Cukurova et al., 2016). Third, emerging research in computer vision and Artificial Intelligence in Education(AIED) show that modern computer vision techniques can be used to detect the behavioural markers we used automatically (Zhou et al., 2023). However, there are still significant challenges in the use of these technologies in real-world contexts so manual work in certain parts of the process is likely to be needed. However, the potential of this research stream to lead to the development of an automatic behaviour detection and feedback provision system is likely which could be greatly benefited from the proposed framework here. Lastly, given the independent, individual sample size and the educational context of this study are limited, further studies with larger sample sizes and different educational contexts are required. Although this work is from a 10-weeks long study, variance between subjects, activities, and context was limited in our sample size. Therefore, more work is needed for the potential generalization and cross-context validity of the proposed approach and our findings.

## 7 Conclusion

This study presents an original method using multimodal learning analytics to detect and analyze educationally meaningful group interactions from video and audio data using machine observable nonverbal speech and gaze behaviours. The identified group interactions are shown to be valuable to generate insights into statistical and process differences in groups with different collaborative learning outcomes. Furthermore, through the lens of process mining, the study provides strategies to support the practice of collaborative learning. However, considering the context dependency of the group interaction status we identified, further explorations of the cross-context validity are needed.

**Acknowledgements** We would like to thank DUTE 2021/2022 students for granting permission to collect data for this study. We also thank Dr Oya Celiktutan (King's College London), Prof. Hajime Nagahara, and Mr Amartya Bhattacharya (Osaka University) for supporting the auto-detection of gaze behaviours with computer vision. At last, we thank UCL Centre for Digital Innovation for providing AWS Doctoral Scholarship in Digital Innovation to the first author, which supports the speaker detection work and the author's PhD study.

**Authors contributions** All authors contributed in constructing the materials, intervention design for the study, data analysis and paper writing. All authors read and approved the final manuscript.

**Funding** None.

**Data availability** The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

## Declarations

**Competing interests** The authors declare that they have no competing interests.



## References

Alwahaby, H., Cukurova, M. (2023). Navigating the ethical landscape of Multimodal Learning Analytics: A guiding framework for research and practitioners. In S. Caballé, J. Casas-Roma, J. Conesa (Eds.), *Ethics in online aibased system*. Elsevier. https://shop.elsevier.com/books/ethics-in-online-ai-based-systems/caballe/978-0-443-18851-0

Alwahaby, H., Cukurova, M., Papamitsiou, Z., & Giannakos, M. (2022). The evidence of impact and ethical considerations of Multimodal Learning Analytics: A systematic literature review. *The Multimodal Learning Analytics Handbook*, 289–325. https://link.springer.com/chapter/10.1007/978-3-031-08076-0_12

Alzahrani, A. S., Tsai, Y., Iqbal, S., Marcos, P. M. M., Scheffel, M., Drachsler, H., Kloos, C. D., Aljohani, N., & Gasevic, D. (2023). Untangling connections between challenges in the adoption of learning analytics in higher education. *Education and Information Technologies, 28*, 4563–4595. https://doi.org/10.1007/s10639-022-11323-x

Amon, M. J., Vrzakova, H., & D'Mello, S. K. (2019). Beyond dyadic coordination: Multimodal behavioral irregularity in triads predicts facets of collaborative problem solving. *Cognitive Science, 43*(10), e12787.

Bohm, D., & Weinberg, R. A. (2004). *On dialogue* (2nd ed.). Routledge. https://doi.org/10.4324/9780203822906

Chen, X., Zou, D., & Xie, H. (2022). A decade of learning analytics: Structural topic modeling based bibliometric analysis. *Education and Information Technologies, 27*, 10517–10561. https://doi.org/10.1007/s10639-022-11046-z

Chua, Y. H. V., Dauwels, J., & Tan, S. C. (2019). Technologies for automated analysis of co-located, real-life, physical learning spaces: Where are we now? *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, 11–20.

Cohen, B. P., & Cohen, E. G. (1991). From groupwork among children to R&D teams: Interdependence, interaction and productivity. *Advances in Group Processes, 8*, 205–225.

Cukurova, M., Avramides, K., Luckin, R., & Mavrikis, M. (2016). Revealing behaviour pattern differences in collaborative problem solving. *Adaptive and Adaptable Learning: 11th European Conference on Technology Enhanced Learning, EC-TEL 2016, Lyon, France, September 13-16, 2016, Proceedings, 11*, 563–569.

Cukurova, M., Luckin, R., Millán, E., & Mavrikis, M. (2018). The NISPI framework: Analysing collaborative problem-solving from students' physical interactions. *Computers & Education, 116*, 93–109. https://doi.org/10.1016/j.compedu.2017.08.007

Cukurova, M., Zhou, Q., Spikol, D., & Landolfi, L. (2020). Modelling collaborative problem-solving competence with transparent learning analytics: Is video data enough? *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 270–275. https://doi.org/10.1145/3375462.3375484

D'angelo, S., & Schneider, B. (2021). Shared gaze visualizations in collaborative interactions: Past, present and future. *Interacting with Computers, 33*(2), 115–133.

Dewiyanti, S., Brand-Gruwel, S., Jochems, W., & Broers, N. J. (2007). Students' experiences with collaborative learning in asynchronous computer-supported collaborative learning environments. *Computers in Human Behavior, 23*(1), 496–514.

Dikker, S., Wan, L., Davidesco, I., Kaggen, L., Oostrik, M., McClintock, J., Rowland, J., Michalareas, G., Van Bavel, J. J., & Ding, M. (2017). Brain-to-brain synchrony tracks real-world dynamic group interactions in the classroom. *Current Biology, 27*(9), 1375–1380.

Dillenbourg, P. (1999). What do you mean by "collaborative learning"? In P. Dillenbourg (Ed.), *Collaborative-learning: Cognitive and computational approaches* (pp. 1–19). Elsevier.

Emery, N. J. (2000). The eyes have it: The neuroethology, function and evolution of social gaze. *Neuroscience & Biobehavioral Reviews, 24*(6), 581–604.

Fan, Y., Saint, J., Singh, S., Jovanovic, J., & Gašević, D. (2021). A learning analytic approach to unveiling self-regulatory processes in learning tactics. *LAK21: 11th International Learning Analytics and Knowledge Conference*, 184–195. https://doi.org/10.1145/3448139.3448211

Gašević, D., Adesope, O., Joksimović, S., & Kovanović, V. (2015). Externally-facilitated regulation scaffolding and role assignment to develop cognitive presence in asynchronous online discussions. *The Internet and Higher Education, 24*, 53–65.

Gašević, D., Joksimović, S., Eagan, B. R., & Shaffer, D. W. (2019). SENS: Network analytics to combine social and cognitive perspectives of collaborative learning. *Computers in Human Behavior, 92*, 562–577. https://doi.org/10.1016/j.chb.2018.07.003

Hadwin, A., & Oshige, M. (2011). Self-regulation, coregulation, and socially shared regulation: Exploring perspectives of social in self-regulated learning theory. *Teachers College Record, 113*(2), 240–264.

Johnson, R. T., Johnson, D. W., & Stanne, M. B. (1985). Effects of cooperative, competitive, and individualistic goal structures on computer-assisted instruction. *Journal of Educational Psychology, 77*(6), 668–677. https://doi.org/10.1037/0022-0663.77.6.668

Kent, C., & Cukurova, M. (2020). Investigating collaboration as a process with theory-driven learning analytics. *Journal of Learning Analytics, 7*(1), 59–71. https://doi.org/10.18608/jla.2020.71.5

Khan, S. M. (2017). Multimodal behavioral analytics in intelligent learning and assessment systems. In A. A. von Davier, M. Zhu, & P. C. Kyllonen (Eds.), *Innovative assessment of collaboration* (pp. 173–184). Springer International Publishing. https://doi.org/10.1007/978-3-319-33261-1_11

Kormanski, C. (1990). Team building patterns of academic groups. *Journal for Specialists in Group Work, 15*(4), 206–214.

Kumar, R., Rosé, C. P., Wang, Y.-C., Joshi, M., & Robinson, A. (2007). Tutorial dialogue as adaptive collaborative learning support. *Frontiers in Artificial Intelligence and Applications, 158*, 383.

Laal, M., & Ghodsi, S. M. (2012). Benefits of collaborative learning. *Procedia-Social and Behavioral Sciences, 31*, 486–490.

Le, H., Janssen, J., & Wubbels, T. (2018). Collaborative learning practices: Teacher and student perceived obstacles to effective student collaboration. *Cambridge Journal of Education, 48*(1), 103–122. https://doi.org/10.1080/0305764X.2016.1259389

Lias, T. E., & Elias, T. (2011). *Learning Analytics: The Definitions, the Processes, and the Potential* (Report). Retrieved from http://learninganalytics.net/LearningAnalyticsDefinitionsProcessesPotential.pdf

Lubold, N., & Pon-Barry, H. (2014). Acoustic-prosodic entrainment and rapport in collaborative learning dialogues. *Proceedings of the 2014 ACM Workshop on Multimodal Learning Analytics Workshop and Grand Challenge*, 5–12. https://doi.org/10.1145/2666633.2666635

Martinez-Maldonado, R., Gaševic, D., Echeverria, V., Fernandez Nieto, G., Swiecki, Z., & Buckingham Shum, S. (2021). What do you mean by collaboration analytics? A conceptual model. *Journal of Learning Analytics, 8*(1), 126–153.

Ouyang, F., Xu, W., & Cukurova, M. (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*.

Ouyang, F., & Xu, W. (2022). The effects of three instructor participatory roles on a small group's collaborative concept mapping. *Journal of Educational Computing Research, 60*(4), 930–959.

Oviatt, S., Hang, K., Zhou, J., Yu, K., & Chen, F. (2018). Dynamic handwriting signal features predict domain expertise. *ACM Transactions on Interactive Intelligent Systems, 8*(3), 18:1-18:21. https://doi.org/10.1145/3213309

Panadero, E., & Järvelä, S. (2015). Socially shared regulation of learning: A review. *European Psychologist, 20*(3), 190–203. https://doi.org/10.1027/1016-9040/a000226

Panitz, T. (1999). *Collaborative versus Cooperative Learning: A Comparison of the Two Concepts Which Will Help Us Understand the Underlying Nature of Interactive Learning*. Retrieved from https://files.eric.ed.gov/fulltext/ED448443.pdf

Pérez Sánchez, C. J., Calle-Alonso, F., & Vega-Rodriguez, M. A. (2022). Learning analytics to predict students' performance: A case study of a neurodidactics-based collaborative learning platform. *Education and Information Technologies, 27*(9), 12913–12938.

Reimann, P., Yacef, K., & Kay, J. (2011). Analyzing collaborative interactions with data mining methods for the benefit of learning. In S. Puntambekar, G. Erkens, & C. Hmelo-Silver (Eds.), *Analyzing interactions in CSCL* (Vol. 12) (pp. 161–185). Springer. https://link.springer.com/chapter/10.1007/978-1-4419-7710-6_8

Schneider, B., & Pea, R. (2013). Real-time mutual gaze perception enhances collaborative learning and collaboration quality. *International Journal of Computer-Supported Collaborative Learning, 88*(4), 375–397. https://link.springer.com/article/10.1007/s11412-013-9181-4

Schneider, B., Worsley, M., & Martinez-Maldonado, R. (2021). Gesture and gaze: Multimodal data in dyadic interactions. In U. Cress, C. Rosé, A. Wise, & J. Oshima (Eds.), *International Handbook of Computer-Supported Collaborative Learning* (pp. 625–641). Springer. https://link.springer.com/chapter/10.1007/978-3-030-65291-3_34

Schoor, C., & Bannert, M. (2012). Exploring regulatory processes during a computer-supported collaborative learning task using process mining. *Computers in Human Behavior, 28*(4), 1321–1331.

Seo, K., Tang, J., Roll, I., Fels, S., & Yoon, D. (2021). The impact of artificial intelligence on learner-instructor interaction in online learning. *International Journal of Educational Technology in Higher Education, 18*, 1–23.

Sharma, K., Olsen, J., Verma, H., Caballero, D., & Jermann, P. (2021). Challenging Joint Visual Attention as a Proxy for Collaborative Performance: ISLS Annual Meeting 2021(virtual): International Society of the Learning Sciences. *International Society of the Learning Sciences, Proceedings*, 91–98. https://doi.org/10.22318/cscl2021.91

Siemens, G., & Baker, R. S. d. (2012). Learning analytics and educational data mining: Towards communication and collaboration. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 252–254.

Slavin, R. E. (1991). Synthesis of research of cooperative learning. *Educational Leadership, 48*(5), 71–82.

Spikol, D., Ruffaldi, E., Dabisias, G., & Cukurova, M. (2018). Supervised machine learning in multimodal learning analytics for estimating success in project-based learning. *Journal of Computer Assisted Learning, 34*(4), 366–377. https://doi.org/10.1111/jcal.12263

Spikol, D., Ruffaldi, E., Landolfi, L., & Cukurova, M. (2017). Estimation of success in collaborative learning based on multimodal learning analytics features. *2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT)*, 269–273. https://doi.org/10.1109/ICALT.2017.122

Stahl, G., & Hakkarainen, K. (2021). Theories of CSCL. In U. Cress, C. Ros´e, A. F. Wise, & J. Oshima (Eds.), *International Handbook of Computer-Supported Collaborative Learning* (pp. 23–43). Springer. https://doi.org/10.1007/978-3-030-65291-3_2

Stahl, G. (2002). Contributions to a theoretical framework for CSCL. *Proceedings of CSCL 2002*. Retrieved from: https://repository.isls.org/bitstream/1/3878/1/62-71.pdf

Sullivan, S., Warner-Hillard, C., Eagan, B., Thompson, R. J., Ruis, A. R., Haines, K., Pugh, C. M., Shaffer, D. W., & Jung, H. S. (2018). Using epistemic network analysis to identify targets for educational interventions in trauma team communication. *Surgery, 163*(4), 938–943. https://doi.org/10.1016/j.surg.2017.11.009

Summers, M., & Volet, S. (2010). Group work does not necessarily equal collaborative learning: Evidence from observations and self-reports. *European Journal of Psychology of Education, 25*(4), 473–492. https://doi.org/10.1007/s10212-010-0026-5

Swing, S. R., & Peterson, P. L. (1982). The relationship of student ability and small-group interaction to student achievement. *American Educational Research Journal, 19*(2), 259–274. https://doi.org/10.3102/00028312019002259

Vinciarelli, A., Pantic, M., & Bourlard, H. (2009). Social signal processing: Survey of an emerging domain. *Image and Vision Computing, 27*(12), 1743–1759. https://doi.org/10.1016/j.imavis.2008.11.007

Vogler, J. S., Schallert, D. L., Jordan, M. E., Song, K., Sanders, A. J., Te Chiang, Y. Y., Lee, J.-E., Park, J. H., & Yu, L.-T. (2017). Life history of a topic in an online discussion: A complex systems theory perspective on how one message attracts class members to create meaning collaboratively. *International Journal of Computer-Supported Collaborative Learning, 12*, 173–194.

Vuopala, E., Hyvönen, P., & Järvelä, S. (2016). Interaction forms in successful collaborative learning in virtual learning environments. *Active Learning in Higher Education, 17*(1), 25–38. https://doi.org/10.1177/1469787415616730

Webb, N. M. (1980). An analysis of group interaction and mathematical errors in heterogeneous ability groups. *British Journal of Educational Psychology, 50*(3), 266–276. https://doi.org/10.1111/j.2044-8279.1980.tb00810.x

Wise, A. F., Knight, S., & Shum, S. B. (2021). Collaborative learning analytics. In U. Cress, C. Rosé, A. F. Wise, & J. Oshima (Eds.), *International handbook of computer-supported collaborative learning* (pp. 425–443). Springer. https://doi.org/10.1007/978-3-030-65291-3_23

Worsley, M., & Blikstein, P. (2011). What's an expert? Using learning analytics to identify emergent markers of expertise through automated speech, sentiment and sketch analysis. In Proceedings of the 4th international conference on educational data mining (pp. 235–239). https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=1dd300f10f22bbebaf3540c35ff1b528a5ea0101#page=247

Zheng, L., Kinshuk, R., Fan, Y., & Long, M. (2023). The impacts of the comprehensive learning analytics approach on learning performance in online collaborative learning. *Education and Information Technologies*. https://doi.org/10.1007/s10639-023-11886-3

Zhou, Q., Suraworachet, W., Pozdniakov, S., Martinez-Maldonado, R., Bartindale, T., Chen, P., Richardson, D., & Cukurova, M. (2021). Investigating students' experiences with collaboration analytics for remote group meetings. In I. Roll, D. McNamara, S. Sosnovsky, R. Luckin, & V. Dimitrova (Eds.), *Artificial intelligence in education* (pp. 472–485). Springer International Publishing. https://doi.org/10.1007/978-3-030-78292-4_38

Zhou, Q., Suraworachet, W., Celiktutan, O., & Cukurova, M. (2022). What does shared understanding in students' face-to-face collaborative learning gaze behaviours "Look Like"? In M. M. Rodrigo, N. Matsuda, A. I. Cristea, & V. Dimitrova (Eds.), *Artificial intelligence in education* (pp. 588–593). Springer International Publishing. https://doi.org/10.1007/978-3-031-11644-5_53

Zhou, Q., Bhattacharya, A., Suraworachet, W., Nagahara, H., & Cukurova, M. (2023). *Automatically detecting gaze behaviours from videos in real-world collaborative learning* (pp. 504–517). Cham: Springer Nature Switzerland. https://link.springer.com/chapter/10.1007/978-3-031-42682-7_34



## Authors and Affiliations

**Qi Zhou[1] · Wannapon Suraworachet[1] · Mutlu Cukurova[1]**

✉ Qi Zhou
qtnvqz3@ucl.ac.uk

Wannapon Suraworachet
wannapon.suraworachet.20@ucl.ac.uk

Mutlu Cukurova
m.cukurova@ucl.ac.uk

[1] UCL Knowledge Lab, Institute of Education, University College London, 23-29 Emerald St, London, UK