CoGE: A Data-Based Framework for Assessing Collaborative Group Engagement in Computer-supported Learning Environments

Carolin Bronowicz[1], Loris Jeitziner[2], Matus Gasparik[1], Valentina Vogel[2], Oliver Rack[2], Lisa Paneth[2], Carmen Zahn[2], Susanne Bleisch[1]

[1]FHNW University of Applied Sciences and Arts Northwestern Switzerland, School of Architecture, Construction and Geomatics, Muttenz, Switzerland

[2]FHNW University of Applied Sciences and Arts Northwestern Switzerland, School of Applied Psychology, Olten, Switzerland

Corresponding author: Carolin Bronowicz; carolin.bronowicz@fhnw.ch

Abstract

To assess collaborative group engagement within computer-supported collaborative learning (CSCL) environments, we introduce a comprehensive, data-based approach, the Collaborative Group Engagement (CoGE) framework. It integrates natural language processing and computer vision in a structured, flowchart-based process. The CoGE framework aims to ensure consistent assessments of behavioral, metacognitive, and socio-emotional dimensions of group engagement. Central to the framework is a temporal segmentation strategy, enabling aligned and contextualized analysis across data dimensions. The time segmentation is used in tailored visualizations of multidimensional data, facilitating nuanced interpretation and informed decision-making through raters. The CoGE framework leverages semi-automation of processes to enhance objectivity and reliability, addressing challenges associated with traditional, subjective assessments of collaborative engagement. The presented framework improves the assessment of group engagement, offering a scalable, objective, and multidimensional approach.

The presented CoGE framework is implemented in a prototype application to test and illustrate its applicability. The application, a dashboard integrates interactive Visual Analytics to support and streamline the interpretation of group engagement, enabling insights, fostering informed data-based interventions, and improving rater efficiency and objectivity. The implemented prototype demonstrates the practical feasibility of the CoGE framework, showcasing its capability to bridge the gap between semi-automated data processing and human-centric evaluation in collaborative learning settings.

Keywords: assessment framework, CSCL, Visual Analytics, nonverbal behavior, collaborative group engagement

1 Collaborative Group Engagement in Computer-Supported Collaborative Learning environments and its current measurement—Relevance for identifying new methods

In today's educational landscape, the concept of computer-supported collaborative learning (CSCL) including AI-support for learning in digital environments is considered a promising approach to enhance learning outcomes in higher education (Crompton & Burke, 2023; Robert, 2024). Such collaborative learning promotes the development of essential skills like problem-solving, creativity and teamwork, which are critical for success in both academic and professional settings (e.g., Ramdani et al., 2022; Trung & Truong, 2023). The integration of digital tools and interactive, intelligent systems into educational practice has created new opportunities to enhance learning (Chen et al., 2018; Niemi et al., 2023). In this vein, understanding the quality of collaborative group engagement (L. T. Jeitziner, Paneth, Rack, & Zahn, 2024; Paneth et al., 2023, 2024; Sinha et al., 2015) is crucial for creating inspiring learning experiences and effective CSCL settings and therefore improving learning outcome.

Collaborative group engagement has been defined by Sinha et al. (2015) as a cognitive and social process. Sinha et al. (2015) outlined four QCGE dimensions in complex digital learning environments: Behavioral Engagement, involving persistence and effort; Social Engagement, marked by respectful, responsive group interactions fostering cohesion and shared understanding; Cognitive Engagement, reflecting collaborative planning, monitoring, and evaluation; and Conceptual Consequential Engagement, where groups apply domain-specific methods, connect ideas to prior knowledge, and understand the broader problem context. For instance, nonverbal cues, such as gestures, facial expressions and body language, which significantly enhance social presence and group cohesion (Burgoon & Dunbar, 2018), were shown to be indicative of collaborative group engagement in both virtual and face-to-face learning groups (L. T. Jeitziner, Paneth, Rack, & Zahn, 2024; Paneth et al., 2024). In consequence, various methods for assessing the quality of collaborative group engagement (QCGE) were developed such as the rating scheme by Sinha et al. (2015) based on verbal communication or the multimethod approach by the Paneth et al. (2023) based on different data sources e.g., verbal and nonverbal data. These methods have proven very promising for the assessment of QCGE (e.g., Paneth et al., 2024). Trained raters are asked to rate several or multiple criteria in a systematic manner. However, the trained raters must rate different criteria simultaneously, which is challenging. The raters must learn to make several complex judgments and decisions based on a complex data set in parallel. For instance, regarding Sinha et al.'s (2015) Social QCGE dimension, a rater must make nine different judgements for different criteria per time segment (see Table 1). According to cognitive load theory (e.g., Van Gog et al., 2011) human cognitive architecture has its limitations regarding parallel information processing of complex data. Thus, the rating task, when it overwhelms the cognitive capacities of the rater, could likely lead to errors, inconsistencies or bias in ratings (L. T. Jeitziner, Paneth, Rack, Zahn, et al., 2024). Automated approaches of assessing collaborative group engagement can reduce the workload by automated data analyses (e.g. through learning analytics) (e.g., Xing et al., 2022). Specifically, in their study, Xing et al. (2022) operationalize social engagement as whether group members have a shorter response time between messages in an online learning environment and whether the messages are reciprocal. Such automated approaches may, however, address a different aspect of engagement than the original construct and potentially lack validity.

A criticism or challenge in the thematic vicinity of collaborative group engagement is the lack of clarity regarding the psychological construct. For instance, the terminology of the four dimensions in Sinha et al.'s (2015) framework for assessing the quality of collaborative group engagement presents some ambiguities. The term "cognitive" may be misleading as it encompasses processes like task monitoring and self-regulation, which are often associated with meta-cognition. Similarly, the social dimension appears to include elements that may align more closely with affective aspects, raising questions about the appropriateness of its label. These ambiguities highlight the need for clearer and more precise terminology in describing the dimensions. The work of Rogat et al. (2022) on Collaborative Group Disciplinary Engagement, which is also based on Sinha et al.'s (2015) work, redefined the "Cognitive" dimension as "Meta-Cognitive" to better reflect the processes it encompasses, such as task monitoring and self-regulation. Additionally, the "Social" dimension was renamed to "Socio-Emotional" to capture the affective components included within its criteria. In sum, existing methods often either burden raters with excessive cognitive load or rely on limited validity of indicators, both of which can alter and distort results. In previous research it was found that ratings based on Sinha et al.'s (2015) methodological approach led to skewed results in the Behavioral and Social dimensions of QCGE (L. T. Jeitziner, Paneth, Rack, & Zahn, 2024; Paneth et al., 2024).

This challenge prompted us to develop a more reliable, valid, objective, and efficient method for assessing collaborative engagement (cf. Hmelo-Silver & Jeong, 2023) which we refer to as the CoGE framework. The framework provides a structured rating system to evaluate the quality of collaborative group engagement in CSCL settings. The framework builds on our previous work (Paneth et al., 2023) and previous defined constructs on collaborative group engagement (Rogat et al., 2022; Sinha et al., 2015). The underlying approach of this framework is based on Visual Analytics for the analysis of high-resolution time series data (Rack et al., 2019) while applying a data-based extension for assessing collaborative group engagement in learning environments. By combining natural language processing and computer vision in a structured flowchart-based process, this framework can contribute to enhance the accuracy and efficiency of collaborative engagement measurement. It preserves the nuanced insights of human judgment while leveraging automated data to support consistent, objective assessments across cognitive and social processes in learning behaviors during group collaboration.

Table 1: Illustration of the challenging parallel judgment requirements of Sinha et al.'s (2015) Social Quality of Collaborative Group Engagement rating scheme.

Social QCGE Criteria	Description	Judgment Required
1. Respectfulness of Conversation	Evaluates respectful communication among group members.	High/Moderate/Low Respectfulness
2. Integration of Contributions	Assesses whether all group members' inputs are acknowledged and integrated.	Full/Partial/Little to No Integration
3. Group Cohesion	Indicates the level of collaborative versus individual work within the group.	Collaborative/Individual/Varied
4. Use of Personal Pronouns	Observes specific language use to identify group inclusivity or individual focus.	Balanced/Inclusive/Exclusive Pronoun Use
5. Group Dominance	Determines if any subgroup or individuals dominate the conversation or decisions.	Balanced/Dominated by Subgroup
6. Conflict Resolution	Assesses the group's ability to resolve tensions effectively.	Fully/Partially/Not Resolved
7. Presence of Competing Ideas	Identifies whether group members present and entertain differing viewpoints.	Multiple Ideas/Single Idea
8. Collaborative Tool Manipulation	Establishes whether group members work together in manipulating shared tools or materials.	Collaborative/Individual Tool Use
9. Disagreement for Further Discussion	Evaluates if disagreements foster constructive dialogue and exploration.	Promotes Discussion/Limits Discussion

1.1 Combination of Visual Analytics and manual rating

The proposed framework builds on the approach outlined by Rack et al. (2019) emphasizing a collaborative research process, structured around two distinct domain areas: traditional group research practices (left) and Visual Analytics (VA) approaches (right) (see Figure 1) (Rack et al., 2019). This setup facilitates the integration of VA methods into established research practices in a way that is flexible and adaptable to the needs of an interdisciplinary research team. VA enhances traditional analytical processes by providing dynamic visual representations of the data that enables simultaneous investigation at different levels of detail and, via transformations, at different abstraction levels. The central element in VA is a closed feedback loop between the computer program executing the data crunching (processing and transforming raw data), generating model outputs and visual representations. The rater, exploring the data by interacting with the visualizations, can form and verify hypotheses by making modifications to the data processing and transformation pipeline, e.g., by changing model or algorithm parameters (Keim et al., 2010; Robinson et al., 2017; Thomas & Cook, 2005). The approach allows for parallel manual coding and exploratory visual analysis. By keeping the analysis open-ended, raters have the freedom to explore the data without constraints or biases toward specific data components, encouraging a creative and flexible investigation that allows patterns, relationships, or insights to emerge naturally. There are problem areas that need to be considered when developing VA tools, as otherwise the rater may misinterpret the visualization of the data, e.g., due to bias or misleading visual representations (Nguyen et al., 2021). In addition, there may also be technical knowledge barriers, as many VA tools require raters to have significant expertise and technical skills, which may hinder accessibility to a wider audience (Khakpour et al., 2023). VA does not aim to replace traditional analysis methods, but extends and enhances them (Rack et al., 2019). For instance, the algorithmic analysis of spatio-temporal movements, such as arm gestures or positional shifts, can guide a more targeted review of selected video segments. Initial coding of group coordination behaviors can also be refined using VA techniques, resulting in a more nuanced understanding of group interactions. Once hypotheses about group coordination are developed, VA designs can be optimized to facilitate the validation of these findings.

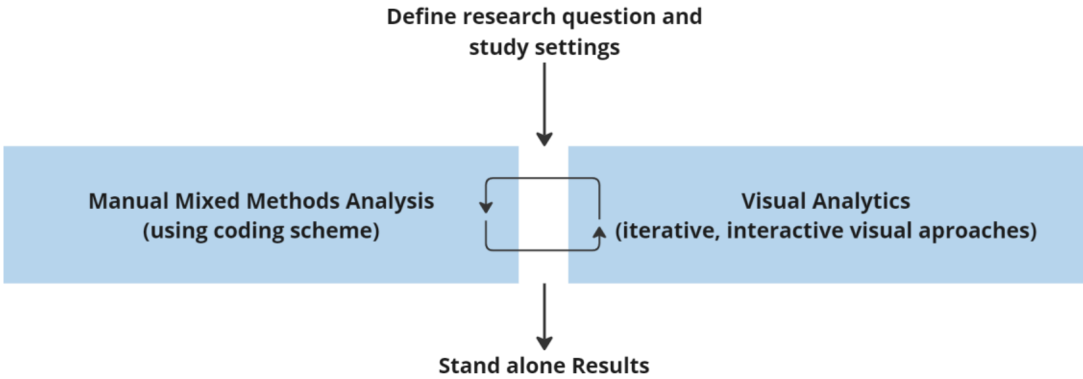

Figure 1: A group research process (left) can be supported by VA approaches (right). Where the iterative VA approaches are integrated with the traditional methods, is defined in an interdisciplinary research team depending on the research questions to be answered. Adapted and simplified from Rack et al. (2019).

We have built our framework based on this idea of combining manual mixed methods analysis with Visual Analytics. Our proposed approach seeks to integrate Visual Analytics as a tool for studying complex multimodal, multi-resolution, quantitative and qualitative data into the established methodological practices of CSCL (group) research in a way that is flexible and adaptable to the needs of specific research objectives.

1.2 Visual Analytics for High-Resolution Time Series

The Visual Analytics (VA) exploration process (see Figure 2), first described by Thomas & Cook (2005), involves a synergistic interaction between automated data analysis and human cognitive strength to identify patterns, designed to facilitate deep insights into complex datasets (Thomas & Cook, 2005). This iterative process typically begins with data ingestion, where large, multi-dimensional datasets are loaded into a VA system. Next, preprocessing techniques such as data cleaning, transformation or aggregation are applied to ensure that the data is accurate and suitable for exploration. Once pre-processed, the data is subjected to automated analysis, which may include techniques such as clustering, pattern recognition or machine learning algorithms that highlight significant trends, correlations and anomalies within the dataset. The visualization phase follows (Keim et al., 2008), where various graphical representations — such as line charts, scatterplots, time-series of density maps, or bar charts — are generated. The choice of visualization depends on the nature of the data and the analytical goals. For example, time-series data might be visualized with line charts to show temporal changes, while multi-dimensional data could benefit from scatterplots or density maps to reveal relationships between variables (Keim et al., 2006). During the exploration phase, raters can dynamically manipulate the visualizations, filtering or zooming in on specific areas of interest by changing or refining the parameters of the models. This iterative process allows raters to explore the dataset from multiple perspectives, refining their inquiries based on visual feedback. VA systems often support exploratory data analysis, where raters can interact with the visualizations in real-time, adjusting parameters such as time intervals or dimensional axes, to discover hidden patterns and refine their hypotheses (Thomas & Cook, 2005). The final stage involves interpretation and decision-making, where insights gained from the visualizations are used to support conclusions and decisions to generate knowledge (Keim et al., 2008). Throughout the repetitive and iterative VA exploration process, the raters ability to interact with the data visually, coupled with automated analysis, creates a powerful feedback loop, enhancing the understanding of complex datasets, uncovering hidden structures and enabling informed decision-making (Deussen et al., 2010; Keim et al., 2006; Kohlhammer et al., 2011; Thomas & Cook, 2005).

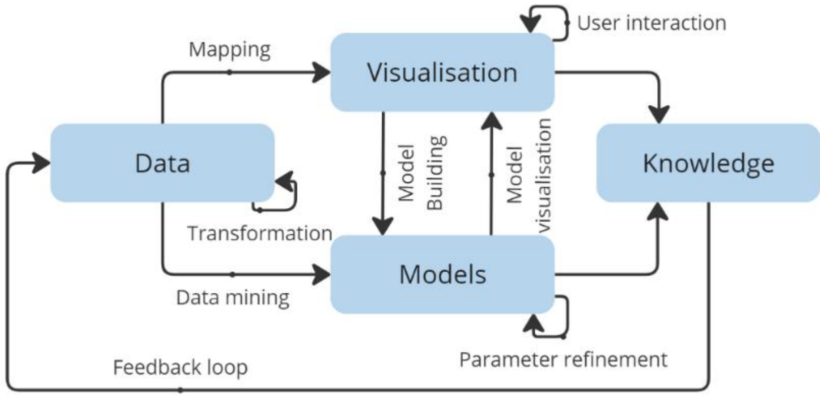

Figure 2: Visual-data exploration: The Visual Analytics process is characterized by the interaction between data, visualizations, models about the data, in an iterative process guided by the raters, to discover knowledge. Adapted from (Keim et al., 2010, p. 10).

In recent years, the application of VA to collaborative learning research has gained traction (see e.g., Hu & Chen, 2021), driven by advancements like tracking methods (e.g., Noroozi et al., 2018) and machine learning (e.g., Schneider & Blikstein, 2015). In group research, tracking methods that bypass cumbersome wearable sensors have been employed alongside machine learning techniques to detect nonverbal patterns, such as body language linked to emotional states (Behoora & Tucker, 2015). For example, Sanchez-Cortes et al. (2012) utilized algorithms to track head movements and other nonverbal behaviors during collaborative problem-solving tasks. Their approach, which included video-recording face-to-face group interactions and automatically extracting nonverbal indicators, enabled early detection of leadership emergence within groups (Sanchez-Cortes et al., 2012). These results highlight the potential of using automated methods to study complex phenomena in real-world settings, such as collaborative learning, as an alternative to or extension of manual data derivation. Parallel advancements in computational methods and interactive visualization technologies have facilitated the development of sophisticated interfaces that enable the concurrent visual analysis of large datasets from multiple perspectives. Modern visualization libraries and tools now allow researchers to integrate machine learning algorithms with visual displays, enhancing the capacity to explore complex data (Endert et al., 2018; Liu et al., 2017). This integration is crucial, as machine learning algorithms often operate as "black boxes," producing results that are difficult to interpret without iterative input by a human domain expert. In this sense, a combination of humans and computers could help to balance the respective strengths and weaknesses. Visual analytics helps bridge this gap by combining computational analysis with human insight, allowing researchers to refine their interpretations and draw better substantiated conclusions. Similarly, Visual Analytics in CSCL research can be used to explore the nonverbal dimensions of QCGE by overlaying video data with behavioral and qualitative analysis. This dual approach allows for a nuanced understanding of how different nonverbal behaviors influence group engagement and collaboration outcomes.

The integration of VA of nonverbal behavior data automatically extracted from video recordings presents new opportunities for advancing CSCL research. These techniques enable a more granular analysis for the rating of group interactions and allow researchers to identify subtle variations in engagement over time. Moreover, by automating aspects of the analysis process, VA can provide real-time insights that could inform interventions, such as those involving social robots, to enhance group collaboration and engagement.

2 CoGE Framework – Combining Visual Analytics and manual rating of group engagement

In this section, we present the CoGE framework for assessing collaborative group engagement in learning groups. This framework is based on automatically processed multi-modal data (audio & visual data) together with a mixed-methods approach (quality ratings and Visual Analytics). Our work builds on prior research regarding the established measurement of collaborative group engagement in computer-supported collaborative learning (L. T. Jeitziner, Paneth, Rack, & Zahn, 2024; Paneth et al., 2023; Rogat et al., 2022; Sinha et al., 2015). However, in this advanced approach, we additionally employ natural language processing and computer vision (CV) (both described in chapter 2.4) to support the qualitative assessment of CoGE. The linguistic markers of QCGE are adapted from our previous work (L. T. Jeitziner, Paneth, Rack, & Zahn, 2024).

2.1 Conceptual structure of the CoGE framework

Our framework structure consists of a criteria-, data-, and rating layer (see Figure 3). At the top is the criteria layer, containing the three defined dimensions including their criteria of collaborative group engagement: Behavioral, Socio-Emotional and Meta-Cognitive. Then in the middle is the indicator layer, which consists of the video layer, the nonverbal layer and the linguistic marker layer. These are arranged horizontally, as they do not have an overarching hierarchy, but can be viewed hierarchically depending on the research question. In addition, several of these layers can be considered together or single layers can be considered repeatedly. The last layer at the bottom is the evaluation layer, which allows CoGE ratings to be obtained.

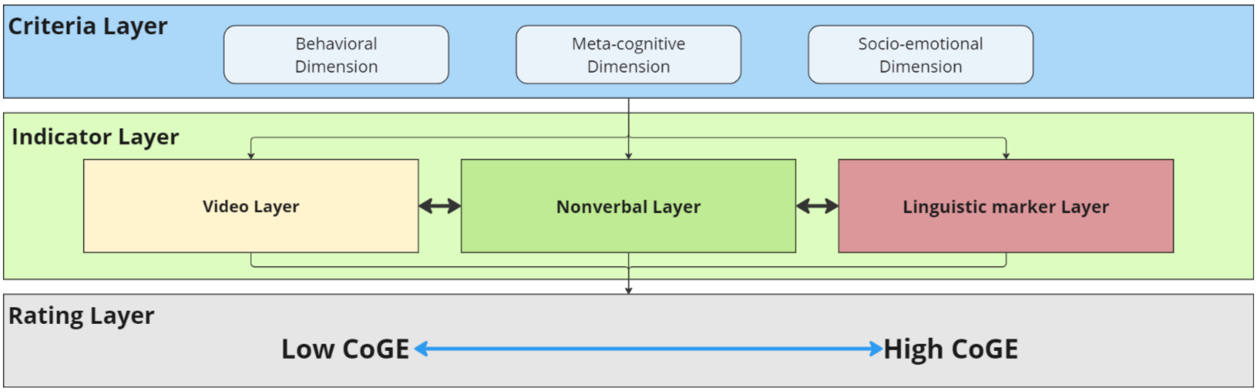

Figure 3: CoGE framework - conceptual structure consisting of three layers: criteria layer, the indicator layer consisting of the video-, linguistic marker- and nonverbal layer, and the rating layer at the bottom.

Based on the presented conceptual structure of the CoGE framework, flowcharts are derived for each criterion of the respective dimension. Figure 6 gives an overview of the CoGE flowchart for the behavioral dimension as one example of all dimensions (for detailed flowcharts see Figure 9 in the supplementary materials). The flowcharts guide the raters in a structured decision-making process.

2.2 Segmentation procedure

Before collaborative group engagement can be assessed using the CoGE framework, the learning sequence of interest must be divided into meaningful parts or segments. Previous research has used rather rigid segmentation procedures, where learning sequences were divided into one- or five-minute units (L. Jeitziner, Paneth, Rack, Zahn, et al., 2024; L. T. Jeitziner, Paneth, Rack, & Zahn, 2024; Sinha et al., 2015). However, the disadvantage of this method is that hard cuts must be applied, which can split an interaction into two segments, although they would naturally be observed and perceived together. For example, if a group is in the middle of a discussion during the one-minute break, this procedure would apply segmentation even though these parts should be observed together. Problems can also arise if outliers (e.g., raising the voice) at the top or bottom are located exactly on the interval boundary. These would then possibly not be correctly included in one of the intervals surrounding the outlier. A lot of relevant information could therefore get lost during the rating process.

We therefore propose a more flexible approach to segmentation but grounded on best practices in the field (Paneth et al., 2024; Sinha et al., 2015). The aim of the segmentation process is to divide a sequence of interest (e.g., the observation of a group completing a task in 60 minutes) into smaller units. Based on our experience, we suggest dividing it into units of between one and two minutes. These time intervals usually contain enough activities of the participants to be able to evaluate them, but at the same time are still manageable and the rater effort is reasonable as some group work and, thus, videos last 1-2 hours. At the beginning, the rater can use the audio track to get an overview of the group discussion. The rater can then use the audio track to identify any pauses that occur naturally. This can be prepared using a dashboard. For example, the audio track can indicate a longer, quiet phase from which the evaluator can set a pause. In this way, longer pauses between utterances can be selected to separate segments. We recommend 1-minute intervals, with the option of moving the limits backwards or forwards by 20 seconds. The following image (Figure 4) shows a segmentation example based on the audio track for a short video section.

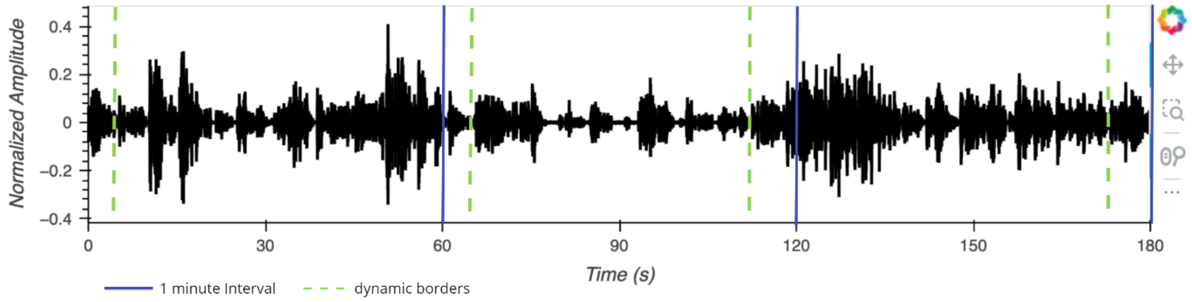

Figure 4: Example of a segmented video section based on the audio track. Aimed interval length of 1 minute with a floating border of +- 20 seconds (green) vs. the standard 1-minute fixed boarders (dark blue)

The segmentation process is shown in detail in Figure 5. Alternatively, the segmentation process can be automated to move the boundary forward or backward by a threshold based on upward and downward outliers.

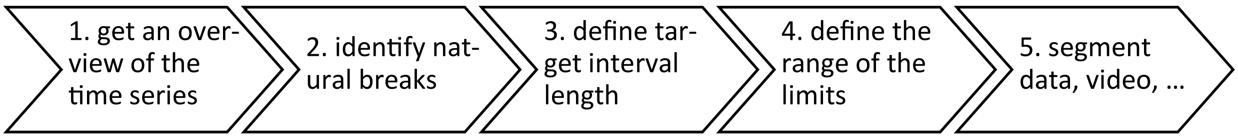

Figure 5: Segmentation process as the base for the video rating

2.3 Criteria Layer

The criteria layer describes the relevant aspects – criteria – of each cognitive group engagement dimension. Each criterion is defined for each dimension and analyzed based on the CoGE flowcharts and rated at the end. The criteria for each dimension and the data that is consulted are described in the following subsections. The list of criteria for each dimension is based on Sinha et al. (2015), whose definitions were adapted to our multi-modal data.

2.3.1 Behavioral Dimension Criteria

The behavioral dimension relates to the following set of important criteria:

Off-/On-task: The Off-/On-task criterium is mainly informed by the content in the group conversation. The rater reads through the transcript of verbal group communication and decides whether the group or some group members are more or less engaged. In addition, the rater can use the verbal amount indicator to judge whether single members are not participating verbally. If the situation is unclear or the rater indecisive, the rater can additionally look at the video recordings or nonverbal layer. For instance, the video may reveal that group members, while not talking, may still be on-task, as could be indicated by nonverbal cues.

Equal collaboration: Regarding equal collaboration, the rater examines whether all group members are on task or if single members are dominating the discourse and others withdraw from it. This is initially informed by using linguistic markers to analyze who is speaking and how frequently, to then calculate an equal collaboration index. Moreover, if the situation is unclear or the rater indecisive, the rater can additionally be informed by the video or nonverbal layer.

Tool usage: Groups can also be on-task without group members talking with each other but instead interacting with tools. Therefore, this criterion can finally inform the group's behavioral dimension of collaborative group engagement. To identify tool usage, the rater first examines whether the groups nonverbal indicators show them facing the direction of the tools (e.g. computers). Mouse tracking or similar analytic tools can additionally provide further information about tool usage. If the rater is indecisive or the situation unclear, the rater may be additionally informed by the video or nonverbal layer.

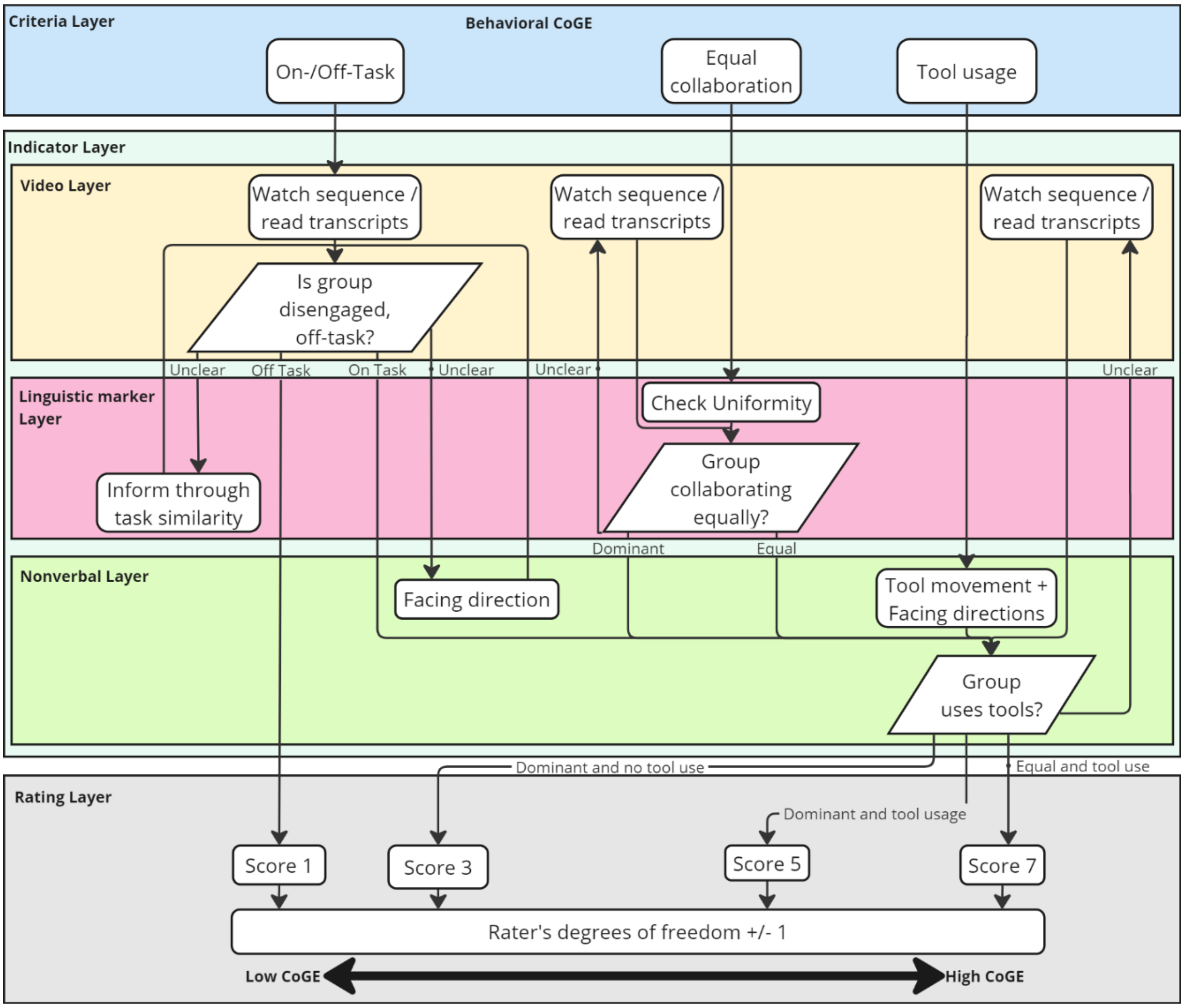

Figure 6: Excerpt of the CoGE flowchart for the behavioral dimension consisting of three criteria "On Task/Off Task", "Equal participation" and "Tool usage" (detailed flowchart of all dimensions and criteria in supplementary materials).

2.3.2 Socio-emotional Dimension Criteria

The socio-emotional dimension relates to the following set of important criteria:

Respectfulness: First, respectfulness in group interaction is assessed. The rater observes a video recorded group interaction sequence or reads the transcript to assess the group's respectfulness in the segment. This step is informed by sentiment analysis (Fischer et al., 2024) of the conversation to detect any negative or disrespectful sentiments. If the group shows disrespectful behavior, the evaluator awards a low score and the evaluation ends, because if the group is disrespectful, it cannot be inclusive or have group awareness at the same time. As this results in a low score without options to improve the score, the rater does not have to assess the other criteria.

Inclusion: The next criterion is inclusion, where the rater examines whether all group members are being included in the conversation and whether their contributions are appreciated. This is also informed by using linguistic markers to analyze who is speaking and how frequently. If the rater determines the group to be exclusive (certain members dominating or marginalizing others and others withdrawing), a lower score is assigned.

Group Mindset: The final criterion in this dimension assesses the group mindset. This involves checking for the usage of first-person pronouns. If members focus more on their individual work rather than the group as a whole, this would signal an individual mindset and a lower score would be assigned. If the rater is indecisive or the situation unclear, the rater may be additionally informed by the video, transcripts or nonverbal layer.

2.3.3 Meta-cognitive Dimension Criteria

The meta-cognitive dimension relates to the following set of important criteria:

Quality of planning: To assess the quality of task and project related planning of the group's collaboration, the rater first gets an overview by watching or reading the segment. Then the rater can judge whether the group approaches their learning task in a structured manner. If the situation is unclear or the rater is indecisive, the linguistic markers time-words can additionally inform.

Monitoring: Regarding monitoring, the rater first gets an overview by watching or reading the segment. Then the rater can judge whether the group monitors their approach to the learning task. If the situation is unclear or the rater is indecisive, the linguistic markers time-words can additionally inform.

Regulation: Regarding monitoring, the rater first gets an overview by watching or reading the segment. Then the rater can judge whether the group regulates themselves. If the situation is unclear or the rater is indecisive, the linguistic markers that indicate whether participants use temporal words (e.g. now, later, after) can additionally inform.

2.4 Indicator Layer

In previous research, verbal (L. T. Jeitziner et al., 2023) and nonverbal (L. T. Jeitziner, Paneth, Rack, & Zahn, 2024; Paneth et al., 2024) indicators of collaborative group engagement were identified. For instance, in segments where learning groups were showing a higher (lower) quality of collaborative group engagement, the group members showed more (less) head nodding, eye contact, gesturing and other nonverbal behavior. Moreover, verbal indicators, or linguistic (conversational) markers were identified that correlate with the quality of collaborative group engagement (Paneth et al., 2024). A higher (lower) amount of conversation or a more (less) equal collaboration in learning groups are linked to a higher (lower) quality of collaborative group engagement (L. T. Jeitziner et al., 2023). Other linguistic markers that can be linked to collaborative group engagement are task similarity, personal pronouns, sentiment density and valence and temporal words (Fischer et al., 2024; L. T. Jeitziner et al., 2023). In addition to manual coding, it is possible to extract nonverbal signals from video recordings without any special equipment using off-the-shelf tools developed for Computer Vision and Augmented Reality applications. For example, Gasparik et al. (in preparation) demonstrated that, by using available algorithms for extracting facial, hand and body landmarks, it is possible to obtain fine-grained body movement data at high temporal resolution. From those, certain nonverbal cues (such as facing direction) can be derived and tracked through time.

Since both nonverbal and verbal indicators can be automatically extracted from audiovisual data, they serve as an additional layer of the CoGE framework (Figure 3). In the CoGE framework, the rater may use the rating criteria and leverage nonverbal and verbal indicators to enrich the assessment of collaborative group engagement based on the observed learning group.

2.5 Rating Layer

The final Rating Layer considers all the dimensions and computes the final score on a scoring range between 0 (low CoGE) and 8 (high CoGE) accordingly. A key component of this layer is the rater's degree of freedom, which allows for a +/- 1 adjustment in the final CoGE score based on the rater's subjective observations. This flexibility ensures the evaluation captures nuances that may not be fully addressed by the semi-automated CoGE flowchart. This allows the rater to make a reflection and refinement round even after the rating has been completed, e.g., to refine indicators that were interpreted differently at the beginning.

3 A dashboard supporting the rating processes

A practical implementation of our framework can be best realized as an interactive dashboard providing visual representations of different data sources. Dashboards have been shown to be effective in integrating and studying multi-modal and multi-dimensional data and streamlining the analytical process (Yigitbasioglu & Velcu, 2012). They also facilitate collaboration between researchers by making all the necessary information accessible in one place (Sarikaya et al., 2019). Designing effective visualization dashboards for large complex data is a non-trivial task. The visualizations need to be tailored to the complexity of the data in order not to introduce a bias (e.g., by inappropriate choice of visual representations) and offer sufficient flexibility in interaction to not constrain the ability of the observer (rater) to explore the data at varying levels of detail and along any dimension. An effective dashboard user interface design should prioritize clarity, simplicity and relevance. The layout must minimize cognitive load by avoiding clutter, using intuitive navigation, and grouping related information (Sweller et al., 2019). Visualization should follow proven principles, such as using appropriate chart types for specific types of data (e.g. bar charts for comparisons, line charts for trends) to facilitate interpretation. Tools such as interactive filtering and drill-down features can improve user control and enable deeper exploration of the data (Sarikaya et al., 2019). Tooltips, explanatory legends, and adaptive features for the color-blind or visually impaired ensure that the interface is inclusive. In addition, performance optimization is critical. Dashboards must load quickly, process large amounts of data efficiently, and perform reliably under a variety of conditions to maintain user confidence and productivity (Burch & Schmid, 2023). Dashboards are powerful tools that, when designed well, improve the efficiency and accuracy of professional workflows. By taking into account design principles, user support, and the inherent limitations of visual data representation, they can transform complex data sets into actionable insights while maintaining the integrity of professional decision-making processes (Yigitbasioglu & Velcu, 2012).

Based on these considerations, a dashboard mock-up was developed (see Figure 7) to make the assessment with the CoGE framework more intuitive, faster and more accurate. The CoGE framework dashboard consists of different parts that are relevant to the framework, so that this can be used for all steps of the rating process in all the layers (cf. Figure 6). This prototype implementation serves to illustrate the integration of various data dimensions. At the current stage, it is not yet employed for performance evaluation or comparative analysis with other tools. Such assessments are planned for subsequent phases of the study. The numbers in the subsequent description of the dashboard parts relate to the numbers in Figure 7 illustrating the parts.

(1) After uploading the video, the rater can set the target segmentation length and the dynamic limit (e.g., one minute and +/- 20 seconds). The segments are given an index number as a name so that the rater does not have to memorize or handle any time stamps. Below the audio track, the rater can see his current ratings for the criteria of each dimension. (2) On the left side are the controls to control the current

segment, to play or pause the video and interactive charts or to skip to the next or last segment. Based on the criteria layer the rater can decide which criteria should be analyzed. Underneath, the rater can tick or un-tick which criteria they wish to assess - this will result in an immediate change to the associated charts and the associated assessment questions listed below. Here the rating layer is incorporated. After the rater has answered the rating questions (yes/no), they can decide whether they want to give an additional point from the degree of freedom of the rating in order to specify the rating result. (3) Then the nonverbal behavior and natural language processing charts are created based on these settings. The charts allow interactions such as linking and swiping to zoom into the data, but the default state is that the entire track is visible for each data dimension, so that the rater sees not only the current segment highlighted, but also the segment in relation to the entire track. The current segment to be evaluated is highlighted in all diagrams to make it easier for the evaluator to establish relationships. The rater can use the suggested charts and data sources to answer the questions or manually change the selection of charts they see on the main screen. If there are questions about the evaluation criteria or the data visualized in the diagrams, the information buttons to the right and left of the headings provide more detailed information. They are hidden so that a trained evaluator only sees the relevant data. (4) At the bottom left, the video for the respective segment and the transcript are played so that the rater can also see what the group talked about. In combination with the diagrams, they can also back up their assessment if they are unsure whether there are any outliers that they cannot understand. This is due to the data layer with their multi-method approach. Finally, the data can be exported so that it can be used for further analyses.

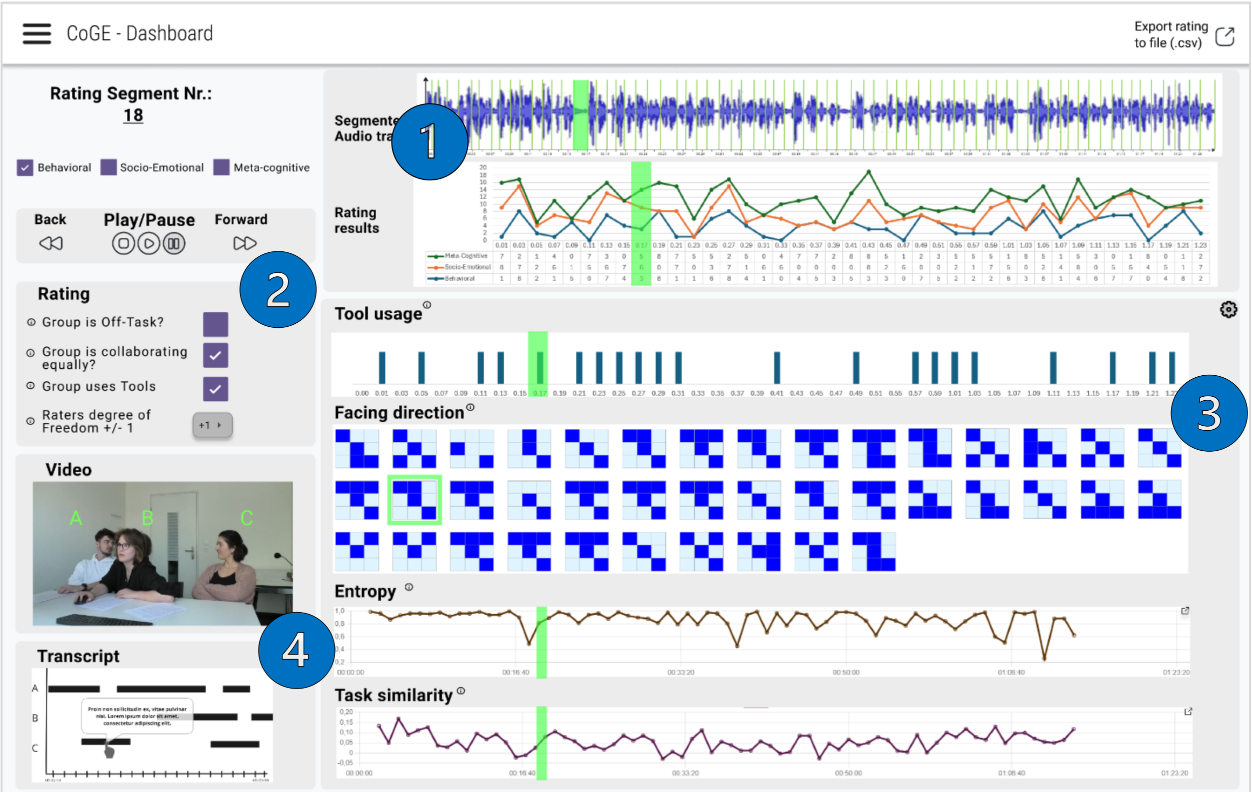

Figure 7: Mock-up for the CoGE framework Dashboard to rate group engagement. (1) data segmentation of audio tracks; (2) segment control and rating section for current segment; (3) interactive, linked visualizations; (4) current video segment and respective transcript

3.1 Prototypical implementation - exemplary use case

To demonstrate how the CoGE framework can be applied to analyzing group work in a CSCL setting, we go through the CoGE flowchart of one dimension including its criteria. This example, which is used to illustrate the framework, originates from a previous study in which three participants completed a learning-by-design task together (Paneth et al., 2024). Communication between all participants was therefore necessary.

A prototype version of the CoGE framework dashboard was used for the analysis, which was developed specifically for this framework and is shown here (Figure 7) in an improved version. The analysis starts by gaining an overview of the data to understand its structure and context. The data was then divided into meaningful segments using dynamic temporal segmentation. For the segmentation of this group, 1-minute intervals were chosen with a possible deviation of 20 seconds. The segmentation was applied to the part of the study to be analyzed. In the next step, the dimension to be assessed was selected, in this case the behavioral dimension of engagement. Relevant charts were activated based on the selected dimension (e.g. equal collaboration, facing direction) and its dimensions, while less relevant charts were deactivated to maintain focus. For the criteria of the behavioral dimension, in addition to the video and the transcript, the diagrams for task similarity, equal collaboration and facing direction were used. Each interval within the dimensions was then assessed using the CoGE flow chart with the help of a binary scoring system. In the case of ambiguities, such as ambiguities in the diagrams, or unexplained outliers, the segments were cross verified with additional resources such as the videos or transcripts to enable an assessment. Finally, the intervals were reviewed to determine if adjustments or corrections were needed, giving the raters flexibility to refine their scores. Here, the diagrams considered were also checked again as a whole to see whether outliers and/or patterns had been overlooked and still needed to be included in the assessment. The CoGE flowchart was then used to create the behavioral CoGE rating for each interval across the three criteria. The same procedure with the corresponding CoGE flowchart would then also be applied to the two other dimensions, socio-emotional and meta-cognitive, to obtain the rating for the observed group across the three dimensions. Figure 8 shows an excerpt of the rating results of the behavioral dimension using the CoGE framework. The first two lines show the segmentation (start and end time of the interval), as well as the difference in seconds to a 1-minute interval. Intervals shorter than one minute are marked in green, intervals longer than one minute are marked in red. The color says nothing about whether this difference is good or bad but only serves to highlight it.

The next block shows the three criteria of the behavioral dimension. The checkboxes show whether the respective criterion in this segment was evaluated as true (checked) or false (unchecked). Below this is the rater's degree of freedom, which is between -1 and +1 and influenced the CoGE rating. This calculation is in the last line, with one rating result per segment.

Behavioral														
Start Time	0:01:00	0:01:50	0:02:50	0:03:30	0:04:20	0:05:20	0:06:20	0:07:20	0:08:20	0:09:20	0:10:10	0:11:20	0:12:40	0:13:50
End Time	0:01:50	0:02:50	0:03:30	0:04:20	0:05:20	0:06:20	0:07:20	0:08:20	0:09:20	0:10:10	0:11:20	0:12:40	0:13:50	0:14:40
Difference to 1 minute intervall	-10	0	-20	-10	0	0	0	0	0	-10	10	20	10	-10
Off-Task	☐	☑	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Uniform	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑	☑
Tool Use	☐	☐	☐	☐	☐	☐	☑	☑	☑	☑	☑	☑	☑	☑
Raters degree of freedom +/- 1	0	1	0	0	0	0	1	1	1	1	1	1	1	1
CoGE Rating	**7**	**2**	**7**	**7**	**7**	**7**	**8**	**8**	**8**	**8**	**8**	**8**	**8**	**8**

Figure 8: Extract from the criteria assessments of the behavioral dimension using the CoGE framework and a dashboard. The first section shows the temporal segmentation of the data. The second section shows the three assessment criteria including the rater's degree of freedom. The last section shows the CoGE rating (between 0 and 8) for each segment calculated based on the decisions of the second section.

4 Discussion

In this paper, we present a novel concept for a data-based framework to assess collaborative group engagement (CoGE) in CSCL groups. This framework is based on a combination of the Visual Analytics process, qualitative observational rating, and automatized methods such as natural language processing and computer vision. Moreover, we adapt and develop existing frameworks for assessing collaborative group engagement (Rogat et al., 2022; Sinha et al., 2015). We present a dashboard mock-up, based on previously employed visualizations and interactions, as a tool for easy application of the CoGE framework.

4.1 Criteria of quality of the CoGE framework

In the following we discuss the CoGE frameworks efficiency, complexity, reliability, objectivity and validity. For example, when a trained rater is instructed to rate behavioral CoGE, they can rely on the linguistic marker "equal collaboration". Subsequently, the rater can use the nonverbal visualization of tool movement and facing direction to judge tool usage. Together with the linguistic marker task similarity and a short summary of the transcripts, the rater can finally judge the behavioral CoGE score of a segment. The CoGE framework aspires to achieve an improvement in the criteria of quality compared to commonly applied observational assessment methods in educational psychology.

4.1.1 Efficiency

The use of visualizations for assessment data accelerates the analysis workflow. The visualization of nonverbal (L. T. Jeitziner, Paneth, Rack, & Zahn, 2024; Paneth et al., 2024) and verbal indicators (L. T. Jeitziner et al., 2023) can be used as a heuristic rating process. Compared to other observational frameworks that relied on a complex assessment process (such as Sinha et al., 2015), we streamline the process through the application of simple flowcharts, rendering the process more efficient while also leveraging the perceptional depth of human observation. In comparison, to rate behavioral quality of collaborative group engagement based on Sinha et al.'s (2015) rating scheme, the rater must examine the full segment of the observation and judge on all the criteria mentioned above simultaneously or alternatively observe and rate the segment at least three times, once for each criterion, therefore rendering the rating process repetitive and laborious.

4.1.2 Complexity

Other observational and automatic frameworks tend to be overly complicated (Sinha et al., 2015) or oversimplifying (Xing et al., 2022) in assessing collaborative group engagement. As illustrated above, in the Sinha et al. (2015) rating scheme multiple parallel judgments need to be made. The high cognitive demand of the rating process may incentivize raters to apply heuristics in the rating process (Kool et al., 2010), e.g., only focusing on a part of the criteria such as equal collaboration, thus reducing the validity of the rated collaborative group engagement. The CoGE framework answers these challenges by guiding the raters to prevent unfavorable cognitive load. While following the CoGE flowchart, raters are asked to judge the criteria sequentially instead of simultaneously. Therefore, one can argue that the CoGE framework can support the complexity of the situation without additional cognitive demand.

4.1.3 Validity

Logical temporal segmentation proves to be a crucial and fundamental step in this process. Segmenting data by time, but also by its natural interruptions (e.g., pauses in speech, speaker changes), ensures that patterns are contextualized, enabling more accurate analyses of the evolution of variables over specific time periods. Moreover, the framework allows raters to check their interpretations against complementary data sets, reducing the likelihood of errors and increasing confidence in their conclusions. In consequence, this framework potentially enhances the validity of assessing collaborative group engagement but also encourages greater engagement with the data, leading to deeper insights and innovative perspectives on collaborative group engagement.

4.1.4 Reliability

Reliability is ensured by the structured process and use of consistent indicators across evaluations. This structured approach minimizes subjective biases by guiding raters through sequential judgments using flowcharts, ensuring consistent application of criteria across different segments and raters.

4.1.5 Objectivity

Objectivity is further enhanced by incorporating visualization techniques that externalize both verbal and nonverbal data (L. Jeitziner, Paneth, Rack, Zahn, et al., 2024; Paneth et al., 2024), reducing reliance on rater intuition. By streamlining the assessment process with simple flowcharts and temporal segmentation, the framework minimizes opportunities for cognitive shortcuts, which could otherwise introduce subjective variability.

4.2 Scientific and practical implications

The CoGE framework has significant practical and scientific implications, bridging the gap between theoretical rigor and real-world applicability in assessing collaborative group engagement. Practically, it streamlines assessment processes through structured workflows and visualizations, enabling educators, researchers, and practitioners to evaluate engagement efficiently and accurately. Customizable dashboards may adapt to diverse needs, supporting applications in education, workplace collaboration, and professional development. Scientifically, the framework lays a foundation for automated analysis tools that leverage machine learning to scale engagement studies, enhancing both the depth and breadth of research. These capabilities make the CoGE framework a transformative tool for refining methodologies, generating actionable insights, and driving innovation in CSCL analysis.

With the help of the presented CoGE framework, we were able to overcome the problems identified in previous studies. The next step includes using the framework in practice and refining it further based on those experiences.

Integrating dashboards into a research framework with multiple data sources provides the ability to hypothesize and uncover additional patterns and relationships in complex data sets. Exploratory capabilities are particularly valuable for identifying new research questions or confirming existing theories. Therefore, the presented dashboard needs to be further improved in terms of its user interface to guide the user but still allowing freedom in deciding on the layout and data source. Furthermore, the potential of Visual Analytics has yet to be fully realized. While the current dashboard provides a foundation for identifying correlations between verbal and nonverbal behaviors, whether linked to group engagement or analyzed independently, future developments could expand its application. The framework also offers a valuable basis for hypothesis generation, interpretation, and categorization, but its effectiveness in uncovering deeper insights will depend on iterative refinement and testing.

4.3 Limitations

This contribution presents certain limitations that must be acknowledged to contextualize its findings and guide future research. As already mentioned, we developed this framework due to the limitations identified in other frameworks and studies. One of the main points was to increase the objectivity of the raters, because in our experience this was not fully guaranteed in these studies. By combining the manual assessment by the rater with the combination of the data and its visualizations, we tried to provide an answer to this problem and were able to test this in the prototype application. However, this objectivity was not tested, which means that we cannot draw any final conclusions as to whether this framework can provide a solution to the objectivity problems identified by the other frameworks. Further research steps are needed, also for broader validation across diverse group engagement use cases.

Our own conceptualization (Paneth et al., 2023, 2024) of collaborative engagement stands in opposition to others. While we accept the nomenclature of Rogat et al.'s (2022) construct, we omit the collaborative and disciplinary components. We argue that the indicator "collaborative" described in their conceptualization is reflected enough in the social dimension. Additionally, we omit the disciplinary dimension, since we do not pursue the contextualization of prior lessons. However, this reduction of the conceptualization of collaborative group engagement may lead to blind spots and therefore a less comprehensive assessment which is a limitation of our work.

When setting up the study environment, it is important to determine whether the CoGE framework should be applied. This will ensure that the setup is suitable for the extraction of nonverbal behavior data, as the correct positioning of participants and cameras is required to capture all relevant information. A non-optimal setting may result in poor quality data related to nonverbal behavior. For example, an unfavorable angle, whereby people are sitting at a large angle to the camera, means that the landmarks cannot be reliably extracted. Another example is when several people cover each other, which also means that no data can be extracted. In this case there would be insufficient data in the rating process. This would be expected to have a negative impact on the rating result. Additionally, the processing of data, including nonverbal behavior and natural language processing, remains a bottleneck due to its time-intensive and semi-automated nature. This introduces the possibility of technical challenges, which could impact the efficiency and scalability of data analysis.

Finally, when conducting studies to assess group engagement, ethical considerations must be made regarding the recording. While these recordings are of relevance to the analysis of behaviors and dynamics, as well as to the CoGE framework itself, they also pose a challenge in terms of participant privacy. It must be ensured that participants are aware of the use and storage of the recordings.

5 Conclusion

The CoGE framework advances the assessment of collaborative group engagement in CSCL environments. By integrating natural language processing and computer vision into a structured, flowchart-based process, the CoGE framework enables a consistent, multidimensional assessment and evaluation of behavioral, metacognitive and socio-emotional dimensions of group engagement. A temporal segmentation strategy forms the basis for aligned and contextual analysis and supports tailored visualizations that facilitate nuanced interpretation and informed decision-making. The CoGE framework addresses the key challenges of traditional engagement assessments by combining human judgement with the objectivity and scalability of semi-automated methods. Its practical feasibility is illustrated with a prototype dashboard implementation. By integrating interactive Visual Analytics, the dashboard provides actionable insights and bridges the gap between semi-automated data processing and human-centered assessment and improves the efficiency and objectivity of evaluators. This work demonstrates the potential of the CoGE framework as a scalable, objective and comprehensive tool for assessing group engagement. Future work will focus on refining the CoGE framework and validating its effectiveness in additional and different settings.

6 Acknowledgements

This work has been funded by the Swiss National Science Foundation within the National Research Program NRP 77 - Project #407740_187258 - Next generation learning: Investigating and enhancing collaborative group engagement quality to support learning groups [by social robots]. For more information see: https://data.snf.ch/grants/grant/187258. We thank all the study participants for their time and engagement. Without them, this work would not have been possible. All data is from participants that have given their informed consent about the use and publication of the collected data including pictures.

All data is from participants that have given their informed consent about the use and publication of the collected data including pictures.

7 References

Behoora, I., & Tucker, C. S. (2015). Machine learning classification of design team members' body language patterns for real time emotional state detection. *Design Studies*, *39*, 100–127. https://doi.org/10.1016/j.destud.2015.04.003

Burch, M., & Schmid, M. (2023). *Dashboard Design*. https://doi.org/10.1201/9781032657301

Burgoon, J. K., & Dunbar, N. E. (2018). Coding Nonverbal Behavior. In E. Brauner, M. Boos, & M. Kolbe (Eds), *The Cambridge Handbook of Group Interaction Analysis* (1st edn, pp. 104–120). Cambridge University Press. https://doi.org/10.1017/9781316286302.007

Chen, J., Wang, M., Kirschner, P. A., & Tsai, C.-C. (2018). The Role of Collaboration, Computer Use, Learning Environments, and Supporting Strategies in CSCL: A Meta-Analysis. *Review of Educational Research*, *88*(6), 799–843. https://doi.org/10.3102/0034654318791584

Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, *20*(1), 22. https://doi.org/10.1186/s41239-023-00392-8

Deussen, O., Ertl, T., & Keim, D. (2010). Visual Analytics. *Informatik-Spektrum*, *33*(6), 549–549. https://doi.org/10.1007/s00287-010-0495-6

Endert, A., Ribarsky, W., Turkay, C., Wong, W., Nabney, I., Blanco, I. D., & Rossi, F. (2018). *The State of the Art in Integrating Machine Learning into Visual Analytics* (arXiv:1802.07954). arXiv. http://arxiv.org/abs/1802.07954

Fischer, O., Jeitziner, L. T., & Wulff, D. U. (2024). Affect in science communication: A data-driven analysis of TED Talks on YouTube. *Humanities and Social Sciences Communications*, *11*(1), 80. https://doi.org/10.1057/s41599-023-02247-z

Hu, L., & Chen, G. (2021). A systematic review of visual representations for analyzing collaborative discourse. *Educational Research Review*, *34*, 100403. https://doi.org/10.1016/j.edurev.2021.100403

Jeitziner, L., Paneth, L., Rack, O., Zahn, C., & Wulff, D. U. (2024). *Automating the Assessment of Collaborative Engagement Using Natural Language Processing*. https://doi.org/10.31234/osf.io/5437v

Jeitziner, L. T., Paneth, L., Rack, O., & Zahn, C. (2024). Beyond words: Investigating non-verbal indicators of collaborative engagement in a virtual synchronous CSCL environment. *Frontiers in Psychology*, *15*, 1347073. https://doi.org/10.3389/fpsyg.2024.1347073

Jeitziner, L. T., Paneth, L., Rack, O., Zahn, C., & Wulff, D. U. (2023). Exploring Linguistic Indicators of Social Collaborative Group Engagement. In C. Damsa, M. Borge, E. Koh, & M. Worsley (Eds), *Proceedings of the 16th International Conference on Computer-Supported Collaborative Learning—CSCL 2023* (pp. 364–365). International Society of the Learning Sciences. https://doi.org/https://doi.org/10.22318/cscl2023.813595

Keim, D., Andrienko, G., Fekete, J.-D., Görg, C., Kohlhammer, J., & Melançon, G. (2008). Visual Analytics: Definition, Process, and Challenges. In A. Kerren, J. T. Stasko, J.-D. Fekete, & C. North (Eds), *Information Visualization* (Vol. 4950, pp. 154–175). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-70956-5_7

Keim, D., Kohlhammer, J., Ellis, G., & Mansmann, F. (2010). *Mastering the Information Age—Solving Problems with Visual Analytics*. Eurographics Association. https://doi.org/10.2312/14803

Keim, D., Mansmann, F., Schneidewind, J., & Ziegler, H. (2006). Challenges in Visual Data Analysis. *Tenth International Conference on Information Visualisation (IV'06)*, 9–16. https://doi.org/10.1109/IV.2006.31

Khakpour, A., Colomo-Palacios, R., Martini, A., & Sánchez-Gordón, M. (2023). The Use of Domain-Specific Languages for Visual Analytics: A Systematic Literature Review. *Technologies*, *11*, 37. https://doi.org/10.3390/technologies11020037

Kohlhammer, J., Keim, D., Pohl, M., Santucci, G., & Andrienko, G. (2011). Solving Problems with Visual Analytics. *Procedia Computer Science, Proceedings of the 2nd European Future Technologies Conference and Exhibition 2011 (FET 11)*, *7*, 117–120. https://doi.org/10.1016/j.procs.2011.12.035

Kool, W., McGuire, J. T., Rosen, Z. B., & Botvinick, M. M. (2010). Decision making and the avoidance of cognitive demand. *Journal of Experimental Psychology: General*, *139*(4), 665.

Liu, S., Wang, X., Liu, M., & Zhu, J. (2017). Towards better analysis of machine learning models: A visual analytics perspective. *Visual Informatics*, *1*(1), 48–56. https://doi.org/10.1016/j.visinf.2017.01.006

Nguyen, V. T., Jung, K., & Gupta, V. (2021). Examining data visualization pitfalls in scientific publications. *Visual Computing for Industry, Biomedicine, and Art*, *4*(1), 27. https://doi.org/10.1186/s42492-021-00092-y

Niemi, H., Pea, R. D., & Lu, Y. (Eds). (2023). *AI in Learning: Designing the Future*. Springer International Publishing. https://doi.org/10.1007/978-3-031-09687-7

Noroozi, F., Corneanu, C. A., Kaminska, D., Sapinski, T., Escalera, S., & Anbarjafari, G. (2018). Survey on Emotional Body Gesture Recognition. *IEEE Transactions on Affective Computing*, *12*(2), 505–523. https://doi.org/10.1109/TAFFC.2018.2874986

Paneth, L., Jeitziner, L. T., Rack, O., Opwis, K., & Zahn, C. (2024). Zooming in: The role of nonverbal behavior in sensing the quality of collaborative group engagement. *International Journal of Computer-Supported Collaborative Learning*, *19*(2), 187–229. https://doi.org/10.1007/s11412-024-09422-7

Paneth, L., Jeitziner, L. T., Rack, O., & Zahn, C. (2023). A Multi-Method Approach to Capture Quality of Collaborative Group Engagement. In C. Damsa, M. Borge, E. Koh, & M. Worsley (Eds), *Proceedings of the 16th International Conference on Computer-Supported Collaborative Learning—CSCL 2023* (pp. 91–98). International Society of the Learning Sciences. https://doi.org/https://doi.org/10.22318/cscl2023.134087

Rack, O., Zahn, C., & Bleisch, S. (2019). Do you see us?—Applied visual analytics for the investigation of group coordination. *Gruppe. Interaktion. Organisation. Zeitschrift Für Angewandte Organisationspsychologie (GIO)*, *50*(1), 53–60. https://doi.org/10.1007/s11612-019-00449-1

Ramdani, D., Susilo, H., Suhadi, S., & Sueb, S. (2022). The Effectiveness of Collaborative Learning on Critical Thinking, Creative Thinking, and Metacognitive Skill Ability: Meta-Analysis on Biological Learning. *European Journal of Educational Research*, *volume–11–2022*(volume–11–issue–3–july–2022), 1607–1628. https://doi.org/10.12973/eu-jer.11.3.1607

Robert, J. (2024). *The Future of AI in Higher Education*. EDUCAUSE. https://www.educause.edu/ecar/research-publications/2024/2024-educause-ai-landscape-study/the-future-of-ai-in-higher-education

Robinson, A. C., Peuquet, D. J., Pezanowski, S., Hardisty, F. A., & Swedberg, B. (2017). Design and evaluation of a geovisual analytics system for uncovering patterns in spatio-temporal event data. *Cartography and Geographic Information Science*, *44*(3), 216–228. https://doi.org/10.1080/15230406.2016.1139467

Rogat, T. K., Hmelo-Silver, C. E., Cheng, B. H., Traynor, A., Adeoye, T. F., Gomoll, A., & Downing, B. K. (2022). A Multidimensional Framework of Collaborative Groups' Disciplinary Engagement. *Frontline Learning Research*, *10*(2), 1–21. https://doi.org/10.14786/flr.v10i2.863

Sanchez-Cortes, D., Aran, O., Mast, M. S., & Gatica-Perez, D. (2012). A Nonverbal Behavior Approach to Identify Emergent Leaders in Small Groups. *IEEE Transactions on Multimedia*, *14*(3), 816–832. https://doi.org/10.1109/TMM.2011.2181941

Sarikaya, A., Correll, M., Bartram, L., Tory, M., & Fisher, D. (2019). What Do We Talk About When We Talk About Dashboards? *IEEE Transactions on Visualization and Computer Graphics*, *25*(1), 682–692. IEEE Transactions on Visualization and Computer Graphics. https://doi.org/10.1109/TVCG.2018.2864903

Schneider, B., & Blikstein, P. (2015). *Unraveling Students' Interaction Around a Tangible Interface Using Multimodal Learning Analytics*. *7*(3), 28.

Sinha, S., Rogat, T. K., Adams-Wiggins, K. R., & Hmelo-Silver, C. E. (2015). Collaborative group engagement in a computer-supported inquiry learning environment. *International Journal of Computer-Supported Collaborative Learning*, *10*(3), 273–307. https://doi.org/10.1007/s11412-015-9218-y

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive Architecture and Instructional Design: 20 Years Later. *Educational Psychology Review*, *31*(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5

Thomas, J. J., & Cook, K. A. (2005). *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. https://ils.unc.edu/courses/2017_fall/inls641_001/books/RD_Agenda_VisualAnalytics.pdf

Trung, D. N., & Truong, D. X. (2023). The benefits of cooperative learning: An overview. *Technium Education and Humanities*, *4*, 78–85. https://doi.org/10.47577/teh.v4i.8709

Van Gog, T., Kester, L., & Paas, F. (2011). Effects of concurrent monitoring on cognitive load and performance as a function of task complexity. *Applied Cognitive Psychology*, *25*(4), 584–587. https://doi.org/10.1002/acp.1726

Xing, W., Zhu, G., Arslan, O., Shim, J., & Popov, V. (2022). Using learning analytics to explore the multifaceted engagement in collaborative learning. *Journal of Computing in Higher Education*. https://doi.org/10.1007/s12528-022-09343-0

Yigitbasioglu, O. M., & Velcu, O. (2012). A review of dashboards in performance management: Implications for design and research. *International Journal of Accounting Information Systems*, *13*(1), 41–59. https://doi.org/10.1016/j.accinf.2011.08.002

8 Supplementary material

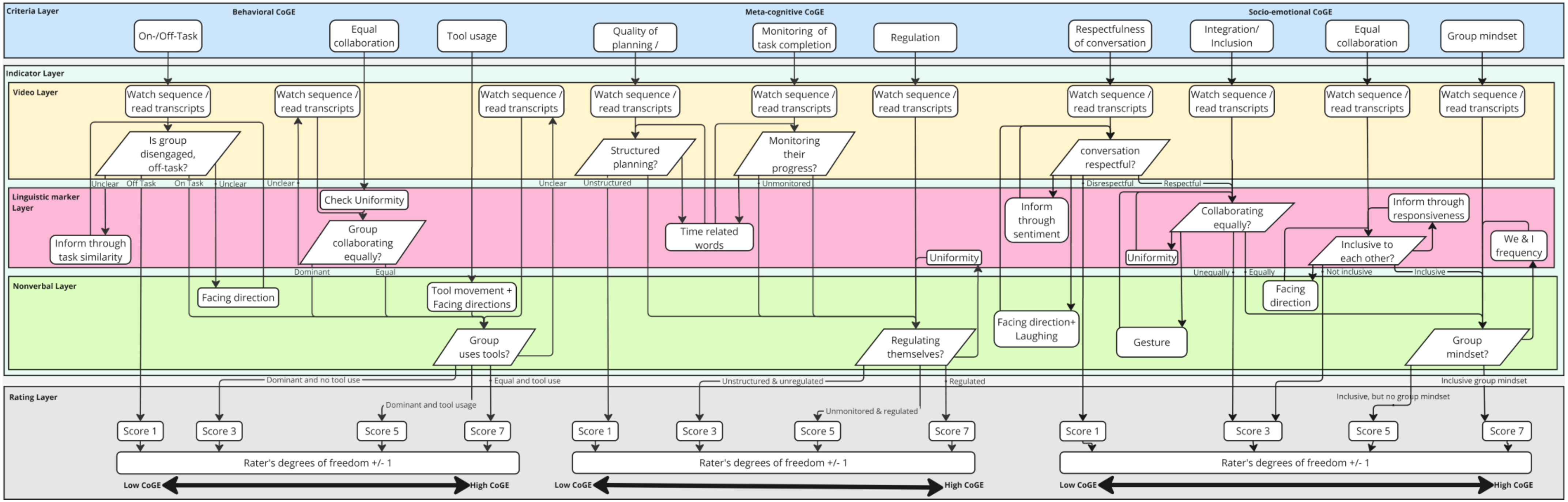

Figure 9: Detailed CoGE flowcharts for all criteria of each dimension