A Multi-Method Approach to Capture Quality of Collaborative Group Engagement

Lisa Paneth, Loris Jeitziner, Oliver Rack and Carmen Zahn

March 28, 2023

A Multi-Method Approach to Capture Quality of Collaborative Group Engagement

Lisa Paneth, Loris T. Jeitziner, Oliver Rack, Carmen Zahn
lisa.paneth@fhnw.ch, loris.jeitziner@fhnw.ch, oliver.rack@fhnw.ch, carmen.zahn@fhnw.ch
University of Applied Sciences and Arts Northwestern Switzerland

Abstract: Multi-method approaches are an emerging trend in CSCL research as they allow to paint a more comprehensive picture of complex group learning processes than using a single method. In this contribution, we combined measures from different data sources to capture the quality of collaborative group engagement (QCGE) in CSCL-groups: QCGE-self-assessments, QCGE-ratings of verbal group communication, and video recorded nonverbal group behaviors. Using different methods of analysis, we visualized, described, and analyzed the data and related the measures to each other. Here, we present results suggesting that measures from different data sources are interrelated: For instance, nonverbal behavior (like nodding the head) is related to high QCGE-ratings of verbal communications. Results are preliminary and show disparities, too. Yet, we conclude that the multi-method approach results in a more comprehensive understanding of QCGE. Feasibility and suitability of the multi-method approach are discussed and conclusions for future research are drawn.

Introduction

Computer-supported collaborative learning (CSCL) has positive effects on skill acquisition, knowledge gain, and student perceptions (Chen et al, 2018). By now, CSCL-research has repeatedly emphasized that it is of utmost importance to understand group learning processes in CSCL-groups *in depth and on different levels* so that they may be appropriately regulated to support successful learning (e.g., Järvelä & Hadwin, 2013; Slof et al., 2020). Our contribution here adds to this research. We present methods developments – in our case: methods to assess the quality of collaborative group engagement (QCGE, Sinha et al., 2015) in CSCL groups – because complex constructs demand appropriate methods to be comprehensively researched and understood, also for improved future options to help groups to regulate their group processes in CSCL. In the next section we will further describe the complex construct QCGE, before we explain our multi-methods approach.

Quality of Collaborative Group Engagement (QCGE) and how to measure it

Sinha et al. (2015) investigated collaborative group engagement in CSCL learning contexts as a construct consisting of four different dimensions on the group level. They postulate four dimensions: 1) *Behavioral engagement* relates to task-related collaboration and defines if group members focus on the common task to be solved (i.e., on-task behavior) or if they rather get distracted (i.e., off-task behavior) and includes on-task persistence and effort investment. 2) *Social engagement* refers to group cohesion and the ways of communicating, whether groups communicate in a constructive and respectful way and whether they are responsive to each other. 3) *Cognitive engagement* addresses the development of a mutual understanding about the common task and of how to proceed in the task as well as the joint regulation of deep-level strategies in group learning. 4) *Conceptual-to-consequential engagement* addresses the ability of learning groups to interrelate diverse content to draw meaningful conclusions, and effectively use technological and conceptual tools to develop group products.

On the empirical level, Sinha et al.`s (2015) four-dimensional conceptualization of engagement and their respective research highlights the complexity and dynamics of the *quality* of those aspects of engagement varying between and within learning groups. The authors could demonstrate in their study using a newly developed video-based observational measure to assess engagement in CSCL-groups and basically quantifying engagement quality according to three levels (1=low, 2=moderate, and 3=high) that groups with low engagement could clearly be differentiated from those with high engagement on the postulated engagement dimensions. Additionally, the authors demonstrated in comparative case studies (by qualitative analyses) how the engagement quality levels on certain dimensions influenced the quality of other dimensions, e.g., high quality behavioral and social engagement fostered high quality cognitive engagement, which then facilitated conceptual-to-consequential engagement. These results are highly important to analyze group engagement in CSCL and point to important options for regulating group processes (e.g., socio-emotional regulation, cf. Hadwin et al., 2017; Järvelä & Hadwin, 2013) and thus how to find ways to support it in CSCL practice in targeted ways.

However, as Sinha et al. (2015) clearly state, further methodological developments are necessary to achieve this goal to investigate the complex nature of group engagement quality more comprehensively and in detail. In the next section we provide a brief overview regarding the scope of CSCL-research methods before we will elaborate on our own multi-methods approach that we have developed to study QCGE.

Methods to capture group processes in CSCL

Seminal CSCL-research has invested important work in developing a range of methods to study group learning processes, e.g., by self- assessment data, which allow to survey motivation, learning and socio-digital activity (Stahl & Hakkarainen, 2021) or by qualitative methods based on different verbal data sources such as verbal conversation analysis (Zemel et al., 2005), content analysis (cf. Trausan-Matu & Slotta, 2021), or dialogue analysis (Hu et al., 2022). Other works additionally included nonverbal behaviors captured in activity transcripts together with verbal communication data (e.g., Barron, 2003; Zahn et al., 2010) or analyzed both transcriptions and gestures from video recordings as a complex process (Trausan-Matu, 2013). Likewise, quantitative approaches based on verbal communication data from audio recordings or written communication in chats were used (e.g., Gijlers et al., 2009; Slof et al., 2016) as well as approaches based on nonverbal behavior data from video recordings, logfiles and data visualizations (e.g., Rack, et al., 2019; Oshima & Hoppe, 2021; Zahn, et al., 2010). Nonverbal body behaviors during social interaction (e.g., gaze, head nodding, gesture) were analyzed in a broad range of research works (from ethnographic research traditions in the Learning Sciences; see Goldmann et al, 2014 to group interaction research on nonverbal expressions of power and dominance in relationships, see Burgoon & Dunbar, 2018). To sum it up, in the field of CSLC and related research, a multitude of methods was applied to understand group processes of learning and motivation.

With growing interest in comprehensively grasping the complexity of collaborative learning processes, there is a related growing interest in combining research methods relying on different data sources (cf. Schneider et al., 2021; Suthers & Lund, 2013), applying computational methods (Oshima & Hoppe, 2021) and including nonverbal behavior. To name just a few: Spikol et al. (2017) detected wrist movements and face orientation in small learning groups using computer vision systems and found that the number of times group members looked at the shared screen and the distance between their hands could identify synchronicity and physical engagement. Grafsgaard et al. (2014) found relations between facial and gestural behaviors and student engagement and frustration in a multimodal study combining self-assessments and automatic recognition of facial and gestural features. In sum, we recognize that in CSCL research multi-modal and multi-method developments have increasingly been appreciated and applied. This is the starting point of our approach presented here with the goal of developing ways to research QCGE as a complex theoretical construct with the group as *the basic unit of analysis* (Stahl, 2015) demanding a multi-method-approach to study it in fair detail. Particularly, we explore and compare *different* methods to examine the *different* dimensions of QCGE (i.e., behavioral, social, cognitive, conceptual-to-consequential engagement quality).

We demonstrate the methods to study QCGE based on (1) self-assessment data from a QCGE-questionnaire (for post-hoc self-ratings in CSCL-groups), (2) verbal communication data and QCGE-ratings on four dimensions (by trained raters), (3) coding of nonverbal group behavior data (by trained coders) and relating them to QCGE-dimensions using visualizations of results to detect possible patterns over the course of task performance in CSCL-groups. We present the initial results from applying this multi-method approach to a data set to demonstrate which insights can be gained from this research practice.

Our main research questions were: <u>RQ1:</u> How can the complex construct QCGE in CSCL-groups be assessed by combinations of measures based on *self-assessment* data, *verbal communication* data, *and nonverbal behavior* data? <u>RQ2</u>: What implications can be drawn concerning interrelations between verbal communication measures of the four dimensions of QCGE and nonverbal behaviors in groups?

Methods

Participants. The sample included a total of N=33 participants (76% female, M= 24.09 Years, SD= 6.70) divided into 11 groups. Participants were students and received either subject hours or financial compensation in return for study participation.

Design of CSCL-Task and QCGE learning context. An authentic CSCL-task paradigm was used for the study which was aligned with related constructionist CSCL research (cf. Harel & Papert, 1991) and design problem solving paradigms (e.g., Goel & Pirolli 1992). Based on McGrath's (1984) task taxonomy, the task paradigm combines a «Planning Task Type 1" and "Creativity Task Type 2" and proved successful for investigating QCGE (Sinha et al., 2015) and CSCL (e.g., "learning by visual design" Zahn, 2017). In the study presented here, the example domain was architectural psychology according to the SNSF-project context of the study (see acknowledgements). Precisely, the task involved groups of 3 students using the 3D modelling tool Sweet Home 3D (eTeks, 2022) to collaboratively design a floorplan for a co-working office according to future users` needs. The task was developed in such a way that individual roles were defined for group members and the respective needs sometimes conflicted with each other. This was deliberately designed to challenge groups on all four dimensions of QCGE. The learning goals of the task include domain-specific knowledge construction (*content-related: architectural psychology)* and collaborative problem solving in groups supported by a complex digital planning tool *(future skills, digital skills)*.

Procedure. Participants received full information about the study and data protection and a written consent form to sign. They were then given 15 minutes to get used to the Sweet Home 3D tool. After this training phase, they worked together for 75 minutes on their self-organized task. To capture verbal communication and nonverbal behavior data, videos were recorded. Finally, after completing the task, participants were given a questionnaire in which they individually rated QCGE in the group.

Table 1
Overview over the measures used in a multi-method approach to capture dimensions of QCGE.

Method	QCGE measure	Data Source	Data Processing	Operationalization & Quantification
(1) Self-assessment	Social Behavioral Cognitive CC	Self-assessment Questionnaire	12 items, 3 items per dimension 7-point Likert-Scale: 1 = no agreement – 7 = total agreement	Descriptive Statistics
(2) Analysis of verbal communication	Social Behavioral Cognitive CC	Audiotrack from video recordings	Transcripts & Expert-rating (trained raters) of 1-minute sequences on each QCGE dimension: Rating: 1= low - 3 = high	Descriptive statistics, Inferential modelling
(3) Analysis of nonverbal behaviors	Computer Operation Activity Hand Gesticulation Propping Head Head Nodding Laughing Bending upper body forward Eye Contact	Video track from video recordings	Observation of displayed behaviors (observable actions) Coding & Counting	Aggregation of the frequencies per 1-minute sequence, Inferential modelling

Note. CC = Conceptual-to-consequential.

QCGE Measurement-Approach: Data collection and data preparation for analysis

An overview of the measures used in our study is provided in *Table 1*. Data were collected from different sources (see Table 1) to receive the following measures: 1) QCGE-self-assessments from a questionnaire (by group members after task completion). 2) QCGE ratings based on verbal communication during collaborative task work (by trained raters). 3) Nonverbal behaviors during collaborative task work (coded by trained coders with a coding scheme). Participants' self-assessments of QCGE dimensions were measured with a 12 items-scale (see Table 2). The scale had partly indicated limited reliability scores ($\alpha < .5$) in a prior validation study, and an improved version is currently being re-validated. Thus, the herein presented analysis and results are taken with caution. Concerning analysis of verbal communication, we used a method related to "quantification of language" (Borge & Rose, 2021) and the QCGE-rating method originally suggested by Sinha et al. (2015, p. 282), but more fine-grained (i.e., based on 1-minute rather than 5-minute CSCL-sequences). Verbal communication from the video recordings was transcribed verbatim and the transcripts were divided into segments corresponding to 1-minute video sequences. Each segment was quality-rated on each of the four QCGE dimensions by two trained raters. We deliberately chose to use transcript based QCGE ratings to ensure a clear separation between verbal and nonverbal communication, thus counteracting potential confounding of the data. With 11 groups completing a task over the course of 75 1-minute sequences, this resulted in 3300 manual ratings. Pre-defined ratings were 1=low, 2=moderate, and 3=high QCGE. Interrater reliability based on *intraclass correlation coefficients* (ICC, cf. Koo & Li, 2016) indicated excellent ICC for all rated QCGE dimensions (ICC = .982, 95% CI [.972, .988], $F(79) = 111$, $p < .001$). Nonverbal behaviors of the participants were collected with a coding scheme that was developed in two steps. Initially, codes were deduced from literature on nonverbal behavioral features indicating emotional states, engagement, and quality of relationships between persons (e.g., Burgoon & Dunbar, 2018) and resulted in nine nonverbal behaviors. The coding scheme was then tested by two trained raters on a 10 minute video of a CSCL-group. Subsequently, two codes were excluded (i.e., "indication by finger" and "mutual gaze"), since the coding of these behaviors from video recordings was too difficult. Interrater reliability based on ICC for each code in the coding scheme revealed good to excellent ICC for all codes (ICC < .602). The final coding scheme consisted of seven codes (see Table 1).

Data Analysis

All statistical analyses were conducted with the statistical software R (R Core Team, 2022). The QCGE self-assessment data were analyzed with descriptive statistics methods. Concerning QCGE ratings based on verbal communication, we used descriptive statistics (e.g., means, standard deviations and visualizations) to explore variance over the course of the task. The ratings over the course of the task were smoothed with *locally estimated scatterplot smoothing* (LOESS). The goal of this approach is to visualize the ratings and detect visual patterns indicating whether there is enough variance for comparisons with the other measures. Concerning nonverbal behavior, we analyzed the frequencies of each code and compared them to the dimensions of QCGE-ratings based on verbal communications. Since the QCGE-ratings indicate ordinal structure of three levels (1=low, 2=moderate and 3=high), we used cumulative link mixed models (Taylor et al., 2022). The nonverbal behavior (see Table 1; column 2) is based on repeated measures of frequencies of nonverbal behavior and the data is hierarchical (i.e., participants are nested in groups), so we applied cumulative link mixed models. Four models were fitted for the four dimensions of QCGE. For each model we defined all nonverbal behavior frequencies as fixed effects. The repeated measure over each one-minute sequence was added as a random intercept. Finally, the participants and groups were applied as random intercepts and the participants were nested in each group to control for individual and group level random effects. We followed the stepwise-elimination approach, proposed by Bates et al. (2015). We ran cumulative link mixed models with the R package ordinal (Christensen, 2019). For all four QCGE dimensions, the random effect of the participant was negligible and we therefore removed the term.

Table 2
Results of the self-assessed QCGE dimensions: behavioral (BE), social (SOC), cognitive (COG), conceptual-to-consequential (CC) engagement.

QCGE dimension	Item	Mean (SD)
BE	All group members actively participated in the task processing.	6.48 (0.67)
	Our group dealt exclusively with the task and nothing else.	6.30 (1.10)
	Our group did not get distracted during the task processing.	6.03 (1.10)
SOC	The communication in our group was respectful and constructive.	6.70 (0.58)
	One or two group members dominated the collaboration and imposed their ideas. *	4.33 (1.78)
	Disagreements sparked constructive discussions.	4.27 (2.21)
COG	Our group worked on the task purposefully.	6.30 (0.81)
	Our group always had the overall task in mind.	5.97 (1.04)
	At the beginning, our group set a concrete plan to solve the task.	5.06 (1.43)
CC	To solve the task, our group linked various pieces of information (from the task sheet and/or prior knowledge).	5.76 (1.11)
	Our group consistently used tools (such as the user manual) to complete the task.	3.42 (1.77)
	The task solution is complete and coherent.	5.12 (1.49)

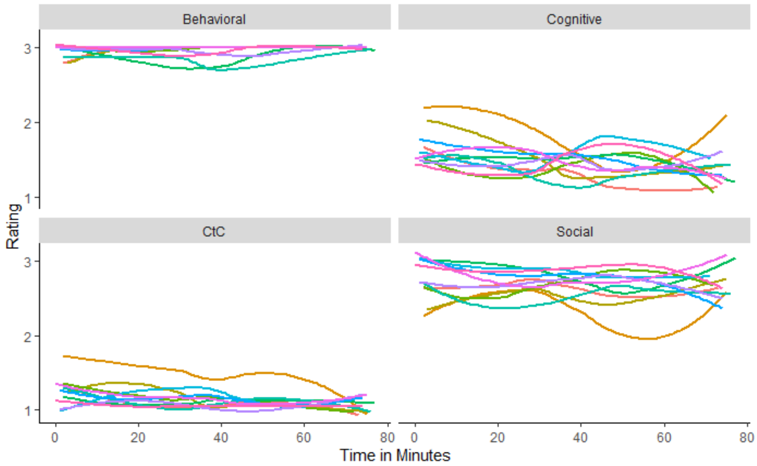
Note. BE = Behavioral, SOC = Social, COG = Cognitive, CC = Conceptual-to-consequential.
* The scale for this item is inverted.

Exemplary Results

We present exemplary results selected for the purpose of providing initial answers to our research questions. We demonstrate how the complex construct QCGE can be assessed based on combinations of measures resulting from self-assessment data, verbal communication data, and nonverbal behavior data (RQ1) and present selected results on interrelations between verbal communication-based measures of the four dimensions of QCGE and nonverbal behaviors in groups (RQ2).

A first set of descriptive results regarding *self-assessed* QCGE (see *Table 2*) indicates that the self-assessments are higher for the behavioral (M = 6.27, SD = 0.99) and cognitive (M = 5.78, SD = 1.71) dimensions of QCGE compared to the dimensions social (M = 5.10, SD = 2.01) and conceptual-to-consequential QCGE (M = 4.77, SD = 1.77). The latter two dimensions also indicate higher standard deviations. We refrained from subsequent (post-hoc) inferential statistical analysis, since the sample size did not meet the requirements.

Figure 1
Smoothed QCGE ratings over the course of the task, categorized by the dimension.



As a second set of results, the QCGE ratings based on *verbal communication* from the transcripts by trained raters, are visualized over the course of the task as illustrated in Figure 1. To visually represent the variance of the ratings over the course of the task per group, the ratings for each group (indicated by different colors) across the 75 minutes of collaborative task work are depicted for all four QCGE dimensions separately - resulting in four visualizations each with 75 ratings of 1-minute sequences *for each group* and *for each QCGE dimension*. The descriptive results suggest that QCGE-ratings on the behavioral (M = 2.96) and social (M = 2.68) QCGE dimensions appear higher than on the cognitive (M = 1.48) and conceptual-to-consequential (M = 1.15) dimension. Moreover, the variance for behavioral QCGE (SD = 0.22) and conceptual-to-consequential QCGE (SD = 0.32) are smaller compared to cognitive (SD = 0.51) and social QCGE (SD = 0.45). This indicates that according to the ratings, the four QCGE dimensions fluctuate in different strengths. Moreover, findings suggest that the groups showed low variance on the dimensions behavioral, and conceptual-to-consequential QCGE.

Table 3
Results of the cumulative link mixed model relating QCGE ratings based on verbal communication to nonverbal behaviors in groups.

QCGE dimension	Nonverbal Behavior	Estimate (OR)	z
Behavioral	Laughing	-0.75 (0.47)	-5.93***
	Eye contact	-0.41 (0.66)	-4.57***
Conceptual-to-consequential	Eye contact	0.16 (1.17)	4.70***
Cognitive	Nodding	0.07 (1.07)	2.08*
	Eye Contact	0.18 (1.20)	6.32***
	Computer operation	0.16 (1.18)	2.10*
Social	Nodding	0.07 (1.07)	0.84+

Note. z = z statistics. OR = Odds Ratio. *** = p < .001, * = p < .05, + = p < .1

A third set of results was calculated from the *nonverbal behavior* data (codings) and concerns interrelations between nonverbal behaviors in CSCL-groups and their QCGE (ratings based on verbal communication). The results of the cumulative link mixed models are presented in detail in Table 3. The models suggest that a number of nonverbal behaviors are significantly associated with the four QCGE dimensions in different ways: 1) Estimates indicate, that for behavioral QCGE, laughing and eye contact between the group members are significantly and negatively associated with the QCGE ratings based on verbal communication data. The more the participants laughed and showed eye contacts with each other, the lower the groups were rated on the dimension of behavioral QCGE. 2) Concerning conceptual-to-consequential QCGE, eye contact is significantly and positively associated. The more the group members showed eye contacts, the higher the groups were rated on the dimension of conceptual-to-consequential QCGE. 3) Concerning cognitive QCGE, it positively and significantly associates with nodding. The more often the group members were nodding, the higher the QCGE rating for the cognitive dimension was. 4) Regarding the dimension of social QCGE, we did not find any significant associations with nonverbal behavior. However, nodding seems to positively associate with the ratings of social QCGE. The estimates (see Table 3) are similar to the model for cognitive QCGE, however, not significant.

Discussion

The goal of this contribution was to find initial answers to two research questions: 1) How can the complex construct QCGE in CSCL-groups be assessed by combinations of measures from *self-assessment data, verbal communication data, and* nonverbal behavior data? 2) What implications can be drawn concerning interrelations between verbal communication measures of the four dimensions of QCGE and nonverbal behaviors in groups?

Concerning the first research question, we provided an empirical example with data from an authentic CSCL-task setting showing how the multi-method approach based on self-assessment data, verbal communication data, and nonverbal behavior data and combinations of measures can complement existing methods (Sinha et al., 2015). We provide example results illustrating the feasibility and which insights can be gained with this approach. Thus, the visualized results from verbal communication based QCGE-ratings illustrate how QCGE fluctuates over the course of the task in an authentic CSCL setting in different strengths implying differences between QCGE dimensions which is in line with previous research by Sinha et al. (2015). The results further indicate that the four dimensions can change independently over time and that it therefore may be possible to design them from the outside in a targeted manner in CSCL-practice. Comparisons between the measures add to the picture: Results from QCGE-self-assessments and from QCGE-ratings based on verbal communication both indicate high levels of behavioral and lower levels of conceptual-to-consequential QCGE compared to the other dimensions across all groups. As for social and cognitive QCGE, the perspectives differ: self-assessments are high for cognitive engagement, whereas QCGE ratings from verbal communication tend to be low to moderate. For social engagement, the picture seems to be reversed - the self-assessments tend to be at a medium level compared to the higher assessed behavioral and cognitive QCGE, whereas the verbal communication based ratings show high values. This finding suggests differences between inside (self-assessment) and outside (trained raters` QCGE-ratings) perspectives and hence, that both solely collecting self-assessments and solely rating groups would each bring about a fragment of the whole picture. However, taken together a more comprehensive view of QCGE emerges, that can be interpreted in more detail and – again could inform CSCL practice. QCGE-self-assessment is thereby important because it represents the learners` perspectives and is as feasible as it gets. The QCGE-rating based on verbal communication over the course of task and coding of the nonverbal behavior is important as an observer's (e.g., future teachers`) perspective. These contrasting perspectives are not the same, as our results show, so research needs both and we suggest to further explore ways to include and combine self-assessments together with observational methods as repeated measure over time.

Regarding the second research question, we found significant relations between verbal communication measures of the four dimensions of QCGE and nonverbal behaviors (Table 3). For instance, higher frequencies of laughing and eye contact relate to lower behavioral QCGE while eye contact seems to be positively related to high scores on the cognitive and conceptual-to-consequential QCGE. Nodding is associated with both cognitive and social QCGE. This indicates that nonverbal behaviors can indeed reflect QCGE dimensions in different ways and thus be used as a further methodological approach to get a clear picture of what QCGE is from interpreting such results. Due to limitations of coding and counting procedures in this regard, combinations with self-assessments or further methods (e.g., reflecting video recordings together with the CSCL-groups) should be considered in the future.

The study has its strengths and limitations: It`s strengths include the relevance of methods development for CSCL-research into QCGE in authentic CSCL-task contexts and the findings we obtained, specifically those showing interrelations between different measures and those concerning nonverbal behaviors. Limitations relate to the limited number of groups (N=11) mitigating general interpretations of results. Concerning the QCGE-ratings based on verbal communication, we also found a limited variance of the ratings aggravating the modelling procedure. QCGE ratings indicated a skew for the social QCGE dimension and similarly, for the conceptual-to-consequential dimension, namely that the groups never showed low social QCGE and never high conceptual-to-consequential QCGE. We presume that this is due to the context of the study and the therein applied task design. Another possible explanation refers, however, to methodological limitations. For instance, those concerning social engagement. Social engagement might have been difficult to assess based on both self-assessments (because they are subject to more complex social processes, such as social desirability) and based on observations (because the QCGE was rated only from verbatim transcripts of the video recordings, but not from the videos themselves). Although this QCGE-rating procedure ensured a clear separation between verbal and nonverbal communication, the lack of context of social interactions when using only the transcripts instead of rich video data may have led to a biased estimation of social engagement. This raises the question of the meaning of 'verbal communication data' and should be considered in follow-up studies and future research. Further on, rating of QCGE and the coding and counting of nonverbal behavior to measure QCGE is laborious and for future analyses, we suggest algorithmizing the rating scheme through methods such as *Natural Language Processing,* which has already been suggested as a promising approach in CSCL-research (Wise et al., 2021). In combination with automatic transcript software, an algorithmic rating of QCGE would render the measurement more feasible and applicable in real-time settings. For nonverbal behavior measures, we consider that recent advances in computer vision and deep learning make it possible to capture videos with ubiquitous consumer-grade hardware (e.g., webcams, smartphone cameras) and to extract 3D landmarks of human faces and body skeletons from them (Lugaresi et al., 2019). This suggests the feasibility of automatically coding nonverbal behavior from video, which is being developed in our research as a method for assessing QCGE. This can, as we hope, further increase the relevance of the respective results for the CSCL Community.

Conclusion

The application and comparison of different methods to capture complex constructs can be a successful way to get a comprehensive picture of their nature. Our results demonstrate how a multi-method approach can be applied successfully to build on and complement existing methods (here: QCGE measures developed by Sinha et al., 2015) in meaningful ways. Our work in progress aims at developing respective computational methods and visual analytics for applied and authentic CSCL-scenarios to measure QCGE. We hope our contribution will stimulate methodological discussions and further research in the CSCL and ISLS communities and in the Learning Sciences.

References

Barron, B. (2003). When Smart Groups Fail. *Journal of the Learning Sciences*, *12*(3), 307–359.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.

Borge, M., & Rosé, C. (2021). Quantitative Approaches to Language in CSCL. In J. N. Lester, C. Rosé, A. F. Wise, & J. Oshima (Eds.), *International Handbook of Computer-Supported Collaborative Learning* (Vol. 19). Springer.

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* (pp. 104–120). Cambridge University Press; Cambridge Core. 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

Christensen, R. H. B. (2019). *Ordinal—Regression Models for Ordinal Data. R package version 2019. 3-9.* http://www.cran.r-project.org/package=ordinal/

eTeks. (2022). *Sweet Home 3D* (7.0). http://www.sweethome3d.com/

Gijlers, H., Saab, N., Van Joolingen, W. R., De Jong, T., & Van Hout-Wolters, B. H. A. M. (2009). Interaction between tool and talk: How instruction and tools support consensus building in collaborative inquiry-learning environments. *Journal of Computer Assisted Learning*, *25*(3), 252–267. https://doi.org/10.1111/j.1365-2729.2008.00302.x

Goel, V., & Pirolli, P. (1992). The structure of Design Problem Spaces. *Cognitive Science*, *16*(3), 395–429.

Goldman, R., Pea, R., Barron, B., & Derry, S. J. (2014). *Video research in the learning sciences*. Routledge.

Grafsgaard, J. F., Wiggins, J. B., Vail, A. K., Boyer, K. E., Wiebe, E. N., & Lester, J. C. (2014). The Additive Value of Multimodal Features for Predicting Engagement, Frustration, and Learning during Tutoring. *Proceedings of the 16th International Conference on Multimodal Interaction*, 42–49.

Hadwin, A., Järvelä, S., & Miller, M. (2017). Self-Regulation, Co-Regulation, and Shared Regulation in Collaborative Learning Environments. In D. H. Schunk & J. A. Greene (Eds.), *Handbook of Self-Regulation of Learning and Performance* (2nd Edition). Routledge.

Harel, I., & Papert, S. (Eds.). (1991). *Constructionism: Research reports and essays, 1985-1990*. Ablex Pub. Corp.

Hu, L., Wu, J., & Chen, G. (2022). iTalk–iSee: A participatory visual learning analytical tool for productive peer talk. *International Journal of Computer-Supported Collaborative Learning*, *17*(3), 397–425.

Järvelä, S., & Hadwin, A. F. (2013). New Frontiers: Regulating Learning in CSCL. *Educational Psychologist*, *48*(1), 25–39. https://doi.org/10.1080/00461520.2012.748006

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, *15*(2), 155–163.

Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., Zhang, F., Chang, C.-L., Yong, M., Lee, J., & others. (2019). Mediapipe: A framework for perceiving and processing reality. *Third Workshop on Computer Vision for AR/VR at IEEE Computer Vision and Pattern Recognition (CVPR)*, *2019*.

McGrath, J. E. (1984). *Groups: Interaction and performance* (Vol. 14). Prentice-Hall Inc.

Oshima, J., & Hoppe, H. U. (2021). Finding meaning in log-file data. In *International Handbook of Computer-Supported Collaborative Learning* (pp. 569-584). Springer, Cham.

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

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

Schneider, B., Worsley, M., & Martinez-Maldonado, R. (2021). Gesture and Gaze: Multimodal Data in Dyadic Interactions. In U. Cress, C. Rosé, A. F. Wise, & J. Oshima (Eds.), *International Handbook of Computer-Supported Collaborative Learning* (pp. 625–641). Springer International Publishing.

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

Slof, B., Leeuwen, A., Janssen, J., & Kirschner, P. A. (2021). Mine, ours, and yours: Whose engagement and prior knowledge affects individual achievement from online collaborative learning? *Journal of Computer Assisted Learning*, *37*(1), 39–50. https://doi.org/10.1111/jcal.12466

Slof, B., Nijdam, D., & Janssen, J. (2016). Do interpersonal skills and interpersonal perceptions predict student learning in CSCL-environments? *Computers & Education*, *97*, 49–60.

Spikol, D., Ruffaldi, E., & Cukurova, M. (2017). *Using multimodal learning analytics to identify aspects of collaboration in project-based learning*. Philadelphia, PA: International Society of the Learning Sciences.

Stahl, G. (2015). A decade of CSCL. *International Journal of Computer-Supported Collaborative Learning*, *10*(4), 337–344. https://doi.org/10.1007/s11412-015-9222-2

Stahl, G., & Hakkarainen, K. (2021). Theories of CSCL. In U. Cress, C. Rosé, A. F. Wise, & J. Oshima (Eds.), *International Handbook of Computer-Supported Collaborative Learning* (Vol. 19). Springer.

Suthers, D., & Lund, K. (2010, June). Productive multivocality in the analysis of collaborative learning. In *Proceedings of the 9th International Conference of the Learning Sciences-Volume 2* (pp. 497-498).

Taylor, J. E., Rousselet, G. A., Scheepers, C., & Sereno, S. C. (2022). Rating norms should be calculated from cumulative link mixed effects models. *Behavior Research Methods*.

Trausan-Matu, S. (2013). Collaborative and differential utterances, pivotal moments, and polyphony. In *Productive multivocality in the analysis of group interactions* (pp. 123-139). Springer, Boston, MA.

Trausan-Matu, S., & Slotta, J. D. (2021). Artifact Analysis. In U. Cress, C. Rosé, A. F. Wise, & J. Oshima (Eds.), *International Handbook of Computer-Supported Collaborative Learning* (Vol. 19). Springer.

Wise, A. F., Knight, S., & Shum, S. B. (2021). Collaborative learning analytics. In *International handbook of computer-supported collaborative learning* (pp. 425–443). Springer.

Zahn, C. (2017). Digital Design and Learning: Cognitive-Constructivist Perspectives. In S. Schwan & U. Cress (Eds.), *The Psychology of Digital Learning: Constructing, Exchanging and Acquiring Knowledge with Digital Media.* (pp. 147–170). Springer International Publishing AG.

Zahn, C., Pea, R., Hesse, F. W., & Rosen, J. (2010). Comparing Simple and Advanced Video Tools as Supports for Complex Collaborative Design Processes. *Journal of the Learning Sciences*, *19*(3), 403–440.

Zemel, A., Xhafa, F., & Stahl, G. (2005). Analyzing the Organization of Collaborative Math Problem-Solving in Online Chats Using Statistics and Conversation Analysis. In H. Fukś, S. Lukosch, & A. C. Salgado (Eds.), *Groupware: Design, Implementation, and Use* (Vol. 3706, pp. 271–283). Springer Berlin Heidelberg. https://doi.org/10.1007/11560296_22

Acknowledgements

This research is funded by the Swiss Science Foundation SNSF within the NRP 77 (Project No. # 187258) and by the University of Applied Sciences and Arts Northwestern Switzerland.