Intern. J. Comput.-Support. Collab. Learn
https://doi.org/10.1007/s11412-025-09445-8

RESEARCH

Measuring the quality of collaborative group engagement: Development and validation of the QCGE self-assessment scale (QCGE-SAS)

Loris T. Jeitziner[1,3] · Lisa Paneth[1,3] · Oliver Rack[1] · Susanne Bleisch[2] · Carmen Zahn[1]

Received: 15 August 2024 / Accepted: 20 March 2025

Abstract
Quality of collaborative group engagement (QCGE) is conceptualised as a complex, shared, and multidimensional construct critical to learning outcomes in collaborative learning. It is currently measured in a variety of ways, including rich observation-based methods. However, a standardised self-assessment method for the construct of QCGE has not yet been developed. This contribution presents a pragmatic self-assessment approach to QCGE for higher education contexts and a new self-assessment scale: the QCGE-SAS. Our findings from a study with $N = 246$ students support the four-factor structure of QCGE originally proposed by Sinha et al. (2015), namely behavioural, social, cognitive, and conceptual-to-consequential collaborative group engagement, and demonstrate that the QCGE-SAS can successfully capture the multidimensional nature of QCGE. Construct and criterion validity were supported by significant correlations with related scales and with self-reported and objective learning outcomes. However, while three subscales proved reliable, the social engagement dimension subscale showed unsatisfactory reliability, indicating further refinement of the scale in future research. Overall, our findings contribute to an accurate operationalisation of the QCGE construct and complement existing methods and comprehensive multi-method approaches to measuring QCGE, thereby informing CSCL research and practice.

Keywords Quality of collaborative group engagement · Self-assessment scale · Subjective measure · Collaborative learning · Higher education

Introduction

In computer-supported collaborative learning (CSCL) in higher education, collaborative group engagement has been identified as a key factor influencing student motivation and, consequently, learning success (Liu et al., 2022). The quality of collaborative group engagement (QCGE) in a CSCL context is conceptualised as shared, contextualised,

Loris T. Jeitziner and Lisa Paneth contributed equally to this work.

Loris T. Jeitziner and Lisa Paneth share the 1st authorship.

Extended author information available on the last page of the article

and multidimensional, that is, distinguishing between the dimensions of behavioural, social, cognitive, and conceptual-to-consequential engagement (Sinha et al., 2015), and has repeatedly been identified as central to collaborative technology-enhanced learning groups (Paneth et al., 2024; Rogat et al., 2022; Sinha et al., 2015).

Despite the growing body of research on QCGE as a complex concept (e.g. Jeitziner et al., 2024a; Rogat et al., 2022), inconsistency remains regarding its operationalisation, particularly in terms of measurement. So far, QCGE has mainly been assessed by observational methods based on comprehensive video analysis procedures (Zahn et al., 2021), for example with qualitative analysis (Paneth et al., 2024; Sinha et al., 2015). However, expert ratings based on observation are not necessarily consistent with learners' self-ratings of group engagement. For example, Paneth et al. (2023) found that student group members rated their own cognitive engagement higher than it was observed (i.e. qualitatively rated); in contrast, they rated their own social engagement lower than it was observed. Thus, a comprehensive measurement approach requires both observation-based and additional introspective measures, such as self-assessments, to include the crucial perspectives of learners on their own group processes when studying CSCL processes in detail. To date, no standardised self-assessment method for a subjective measurement of QCGE has yet been developed. Existing scales that assess constructs and characteristics similar to QCGE often take the individual student as the unit of analysis (i.e. engagement is not conceptualised as shared) and focus on different aspects of collaboration other than QCGE.

In addition to the methodological need for better measurement of QCGE, there is a practical need for and benefit to a standardised self-assessment method for QCGE: in applied contexts, such as university teaching, self-assessment methods for collaborative group engagement can add value to group assessment, e.g. formative evaluations at group level. They can provide lecturers with information about the quality of collaborative work in learning groups, thus emphasising group work in computer-supported higher education settings and promoting important future skills (Forsell et al., 2020; Tumpa et al., 2022). In terms of supporting learning groups in practice, a QCGE self-assessment could be used as a basis for adapting and developing learning processes within learning groups and could usefully be integrated into group-awareness tools (Chen et al., 2024; Schnaubert & Bodemer, 2022) or automated group engagement feedback tools for student groups (Zheng et al., 2023a, 2023b). This can provide student groups with information about the quality of their own group engagement, and thus, an important opportunity to reflect on and regulate their group processes, which in turn can lead to better group learning outcomes (e.g. Zheng et al., 2023a, 2023b).

On the basis of the research desideratum for a self-assessment scale to measure QCGE and the high relevance of self-assessment in higher education learning groups, this paper aims to present a standardised QCGE self-assessment scale (QCGE-SAS) that we developed and validated in our study ($N = 246$). The main goal is to provide a new and pragmatic self-assessment scale that can be easily used and evaluated in CSCL research and practice in higher education contexts, compared with – and complementary to – existing complex observation-based procedures (Paneth et al., 2024).

With this study we aim to answer the following research question:

> How can the quality of collaborative group engagement be measured using a self-assessment scale measuring the four dimensions: behavioural, social, cognitive, and conceptual-to-consequential engagement in higher education contexts?

In the following sections, we present a theoretical background before outlining our methods, findings and discussion of the findings. We conclude with scientific and practical implications.

Theoretical background

Quality of Collaborative Group Engagement (QCGE)

Collaborative group engagement, viewed as a multidimensional core process that links motivation, effort, and learning success at the group level (Sinha et al., 2015) is a key element for the shared regulation of learning in CSCL (Järvelä & Hadwin, 2013). Yet, the conceptualization of engagement in a CSCL context, i.e. at the group level, has only recently been systematically explored (e.g. Paneth et al., 2024; Rogat et al., 2022; Sinha et al., 2015; Xing et al., 2022). Beyond these frameworks, other studies have emphasized the dynamic and interconnected nature of individual engagement dimensions. For instance, Fredricks et al. (2004) proposed a tripartite model of engagement encompassing behavioral, emotional, and cognitive aspects, which has been widely influential in educational research. Although this model is more general, its emphasis on the interplay between dimensions aligns with the notion that high-quality collaborative engagement arises from the mutual reinforcement of persistence, positive interpersonal dynamics, and deep cognitive processing.

On the group level, a sound theoretical conceptualization has been provided by Sinha et al. (2015). The authors identified four dimensions of QCGE in complex digital learning environments: behavioural engagement (BE) includes persistence and effort in the task; social engagement (SE) involves the quality of group interactions, manifested by respectful, responsive exchanges, whereby high-quality SE facilitates shared understanding among group members and promotes group cohesion; cognitive engagement (CE) includes the collaborative development of understanding, which is reflected in task-centred monitoring, planning and evaluation; and conceptual-to-consequential engagement (CC) is defined as student groups using domain-specific methods to solve problems, activating solutions, connecting ideas to prior knowledge and understanding the broader context of the problem. Sinha et al. (2015) tested their framework of QCGE with ten groups of students and an observational approach. They found that the four dimensions are interrelated: for example, BE and SE are important prerequisites for building high levels of CE, which in turn facilitates CC. They also found that learning groups with higher quality of overall engagement had better learning outcomes. Rogat et al. (2022) built on this framework and further developed it for science, technology, engineering and mathematics (STEM) education and related disciplinary practices. They also used an observational approach and found that high quality in one dimension of QCGE was related to high quality in other dimensions. Paneth et al. (2023), using a mixed methods approach with both observations and self-assessments, found that subjective and objective assessments of QCGE differed: for example, students rated their own cognitive engagement as higher and their social engagement as lower than expert raters did. Jeitziner et al. (2024b) hypothesized that linguistic indicators such as sentiment, semantic similarity to task instructions, or verbal mimicry could predict QCGE using natural language processing (NLP). In addition, using an observational approach and qualitative case studies, Paneth et al. (2024) found that nonverbal group behaviours were related to different dimensions of QCGE and that the dimensions of QCGE did not *always* have to be high for the collaboration to be productive: for example,

sequences of low-quality behavioural engagement (e.g. off-task joking) tend to support group cohesion and thus social engagement.

Research also suggests that high-quality collaborative group engagement can impact learning outcomes. Sinha et al. (2015) found that higher levels of engagement correspond to better learning outcomes. Liu et al. (2022) found positive effects of group engagement on learning outcomes in massive open online course (MOOC) discussions, and Curşeu et al. (2020) found that group learning engagement had a positive impact on group performance. Mixed results were found by Xing et al. (2022): in their research, they found that behavioural and cognitive engagement had a positive effect on group problem solving performance, whereas social engagement had a negative effect. Conversely, the results of our own work (Paneth et al., 2024) showed no correlations between QCGE on different dimensions and group product quality in a CSCL task.

To conclude, despite the growing body of research on collaborative group engagement, operationalisations of this concept remain inconsistent. Therefore, scholars emphasize that understanding engagement in learning, especially in collaborative contexts, requires further research, using new and different methods to examine it from different perspectives (e.g. Paneth et al., 2024; Rogat et al., 2022; Xing et al., 2022).

Methodological perspectives and approaches in CSCL

Research in computer-supported collaborative learning (CSCL) has employed diverse methodological approaches to explore the dynamics and outcomes of collaborative learning. Early foundational studies, such as those by Meier et al. (2007), emphasized the importance of structured observation and coding schemes to assess collaborative processes. Their work demonstrated that group performance and interaction quality could be systematically analysed through behavioral indicators and task-oriented metrics. Similarly, Détienne et al. (2012) extended this approach by focusing on the interplay between task-oriented and socio-emotional dimensions of collaboration. Their contributions highlighted the multifaceted nature of group engagement and the need for methodologies that capture both explicit interactions and the underlying social processes driving collaboration.

These seminal contributions laid the groundwork for a variety of methods in CSCL research, including video-based analysis, qualitative coding, and mixed-methods approaches. Observation-based techniques, while providing rich data on group interactions, are resource-intensive and often limited in capturing learners' subjective experiences. To address this gap, researchers have increasingly turned to self-assessment tools, which allow participants to reflect on and report their perceptions of collaborative processes. Such tools provide valuable insights into the subjective and intersubjective dimensions of learning, complementing the objective data obtained from observation-based methods (Hmelo-Silver & Jeong, 2021; Paneth et al., 2023).

A further theoretical distinction of different perspectives on the measurement of learning in CSCL is made by Stahl and Hakkarainen (2021). The authors distinguish between a *subjective*, an *intersubjective*, and an *inter-objective* focus of collaboration in CSCL. While subjective CSCL perspectives acknowledge that students' learning is significantly influenced by the social context, they measure the effects of collaboration on the individual. Intersubjective perspectives, on the other hand, focus on the group as the unit of analysis, where collaborative learning has an impact on other levels, such as individual learning and that of larger groups or communities. The inter-objective perspective focuses on communal or cultural levels of analysis and emphasizes the interaction of learning groups with

networks and artifacts. In this article, we will focus on the *subjective* as well as the *inter-subjective* perspective, i.e. the assessment of group level engagement by individual learning group members, based on their perception of collaboration within the learning group.

Thus, a variety of methodological perspectives and approaches have been developed by CSCL scholars to study different aspects of group learning processes, including quantitative and qualitative approaches such as video observations and analysis of learning groups (cf. Zahn et al., 2021), rating and coding of verbal and nonverbal group communication (e.g. Paneth et al., 2024; Rack et al., 2018), and self-assessments to measure motivation, perceptions and socio-digital activity of learning groups (Hmelo-Silver & Jeong, 2021). A distinction can be made here between *observation-based* and *subjective* research measures. Observation-based measures, such as video recordings or physiological responses, allow researchers to quantify and evaluate both conscious and unconscious processes of the observed study participants, whereas subjective measures focus on evaluating the conscious identification and evaluation of participants' perceptions (Ferreira & Saraiva, 2019). In addition, subjective methods of measurement, such as self-assessments, take a student-centred approach that is central to a detailed assessment of students' attitudes and emotional states toward learning (Paneth et al., 2023). Both approaches are meaningful and provide an important and unique perspective on learning processes. As emphasized in recent literature (e.g. Paneth et al., 2023; Pekrun, 2020; Vriesema & McCaslin, 2020), the combination of observational and self-assessment measures within multi-method approaches is important to provide a more holistic view of these learning processes.

Self-assessment scales for collaborative engagement

Within the learning sciences, many self-assessment questionnaires and scales have been developed to measure motivational aspects of learning, and specifically, student engagement [for a comprehensive review of student engagement scales in higher education, see Heilporn et al. (2024)]. For example, there are various self-assessment questionnaires of student engagement in higher education, such as the well-known Indiana University *National Survey of Student Engagement* (NSSE, 2016), or a variety of scales to collect self-assessment data on specific dimensions of student engagement, such as the *Cognitive Engagement* subscale as part of the *Motivation and Strategy Use Survey* (MSUS) developed by Greene and Miller (1996). To our knowledge, there are no comprehensive self-assessment scales that measure the self-reported quality of collaborative group engagement in a CSCL environment conceptualised as described above. However, there are several questionnaires and scales that assess self-reported group-level engagement within similar frameworks or for similar aspects of collaboration in learning. In the following, we describe and compare relevant publications.

For example, the MSUS (Greene & Miller, 1996) scale measures the level of cognitive engagement of students during their studies. It consists of two subdimensions: meaningful cognitive engagement (reflects deep processing and self-regulatory activities that students engage in while studying, sample item: 'I made a plan for achieving the grade I wanted on this exam.'); and shallow cognitive engagement (surface-level processing and memorization strategies used by students, sample item: 'I tried to write down exactly what my instructor said during lectures'). The scale consists of 25 items for meaningful cognitive engagement and 13 items for shallow cognitive engagement that are rated on a Likert-type scale.

The *Pair Programming Mutual Engagement (PPME) Questionnaire* was developed by Xu and Correia (2024). It was originally developed for middle school student dyads in a computer science course who were engaged in pair programming. The authors used Sinha et al.'s (2015) QCGE framework (i.e. for the dimensions of behavioural and cognitive engagement), among other relevant literature, to develop the items. The PPME questionnaire consists of 20 items and four subscales, namely behavioural engagement, cognitive engagement, emotional engagement and social engagement. Sample items are: 'I contributed to our project' for behavioural; 'we set specific goals for completing the project' for cognitive; 'I looked forward to today's activity' for emotional; and 'I feel my opinions have been taken into account' for social engagement.

In addition, the *Collaborative Learning Engagement Scale* was developed and validated for Chinese undergraduate and master's students by Xu et al. (2024). The authors employed a three-factor model of student engagement (behavioural, cognitive and emotional) using items from well-established measures of student engagement in individual, non-group contexts. The scale consists of 29 items and the three subscales mentioned above. Sample items are: 'I actively interact with peers in my group' for behavioural engagement; 'I make enough effort to do our group work' for cognitive engagement; and 'when I think about our group work, I feel worried/relaxed' for emotional engagement.

The *group regulation scale* was developed by Lai (2021). The scale has a 5-point Likert rating and consists of 12 items and four sub-dimensions, namely effort regulation, goal expectancy, help-seeking and time management. Sample items are as follows: 'in this lesson, we try to do the tasks given to us by the teacher' for effort regulation; 'we think that we have paid enough attention to this lesson' for goal expectancy; 'if we have problems with the project, we ask the trainer for help' for help-seeking; and 'we try to plan at least once a week to work together on the project' for time management.

Ito and Umemoto (2022) developed a scale to assess *self-, co-, and socially shared regulation of intrinsic motivation* for collaborative activities. This scale consists of nine items and three subscales: self-regulation, co-regulation and social regulation of intrinsic motivation. The scale aims to identify how intrinsic motivation affects the three modes of regulation in relation to experiences of group activities in higher education tasks. Sample items are as follows: 'I try to increase my motivation by making the contents of group activities as enjoyable as possible' for self-regulation; 'I try to support the motivation of the group member by making the contents of group activities as enjoyable as possible' for co-regulation; and 'I try to support the motivation of the entire group by devising ways to make them find the contents of group activities interesting' for socially-shared regulation of intrinsic motivation.

Biasutti and Frate (2018), for their part, developed a *group metacognition scale* for university students who participated in an online group study. The scale is composed of four sub-dimensions (knowledge of cognition, planning, monitoring and evaluation) and 20 items. The items of the scale provide a 5-point Likert type rating as 'strongly disagree', 'disagree', 'undecided', 'agree', and 'strongly agree'. Sample items are as follows: 'we know our strengths as learners' for knowledge of cognition; 'we ask questions to check our understanding' for monitoring; 'we make judgments about workload' for evaluation; and 'we select the appropriate tools' for planning. Another scale for measuring metacognition in online learning groups was developed by Zheng et al., (2023a, 2023b), namely the Social Metacognition Inventory with 24 indicators assessing beliefs about others, awareness of others' thinking, judgment of others' emotions, co-regulation of others' thinking and evaluation of others' thinking.

Further scale development for constructs similar to QCGE has been done, for example, by Atxurra (2015, *Cooperative Learning Application Scale*), Xiong et al., (2015, *The Learners' Perceived Readiness for Computer-Supported Collaborative Learning* questionnaire), and Bojanova (2013, *Scale for Student Engagement in Group Projects*).

Research on engagement has not only been conducted on learning groups but also on work teams. In developing the *QCGE self-assessment scale* (QCGE-SAS), we also draw on established instruments in this area, recognising the significant similarities between learning groups and work teams (Prichard et al., 2006). For example, the *Questionnaire on Teamwork* (FAT) is a well-established questionnaire for measuring the behaviour of working teams such as project teams, interdisciplinary and semi-autonomous working teams, developed by Kauffeld and Frieling (2006). It is based on two scales and respective subscales: (1) social reflexivity – with the subscales of cohesion (example item: 'the team is in the centre and not the individual') and assumption of responsibility (example item: 'everyone contributes equally to the team'); and (2) task reflexivity – with the subscales of goal orientation (example item: 'our goals are realistic and achievable') and task mastery (example item: 'our priorities are clear'). In terms of the QCGE construct, the first scale (social reflexivity) is particularly similar to the social engagement dimension and therefore highly relevant for our purpose. In addition, team engagement has been extensively studied by Costa et al. (2014) and is referred to as *Team Work Engagement* (TWE). The authors derived the TWE scale from the Utrecht Work Engagement Scale (UWES, Schaufeli et al., 2006), which measures individual work engagement, and linguistically adapted and validated it using the first-person plural ('we'). The scale consists of nine items and three dimensions, namely: vigour (sample item: 'at our work, we feel bursting with energy'); dedication (sample item: 'we are proud of the work that we do'); and absorption (sample item: 'we get carried away when we are working').

In developing the QCGE-SAS, we draw on established instruments in this area. We also build on our prior research and validation efforts related to this scale. In a previous study (Paneth et al., 2023), we examined a preliminary, unvalidated version of the QCGE-SAS. While the results of this previous study provided valuable insights, they revealed certain limitations, including ceiling effects and low reliability in measuring the social dimensions of the scale.

Table 1 lists the above-described selection of scales and questionnaires that measure aspects, dimensions or constructs that we have found to be similar to the concept of QCGE. Table 1 includes the names of the scales, the authors, the participants or target group for whom the scales were developed, the theoretical or empirical background of the scale development, the content (number of dimensions/items), as well as sample items. The focus of Table 1 is on self-assessment in collaborative learning, as we developed the QCGE-SAS for a CSCL context.

In summary, the scales and questionnaires presented and summarized in Table 1 measure characteristics similar to the QCGE construct. However, many of them take the individual student as the unit of analysis and, in addition, no comprehensive scale has been developed that conceptualises the QCGE construct in a CSCL context in higher education *and* across the four dimensions of behavioural, social, cognitive, and conceptual-to-consequential engagement. We argue that in line with the conceptualisation of QCGE as a multidimensional construct focused on CSCL environments, the scale developed in this study should consider the group as the unit of analysis and phrase the items accordingly (Stahl & Hakkarainen, 2021).

Given the theoretical elaboration of the concept of QCGE (Sinha et al., 2015) and its close grounding in further relevant research (e.g. Paneth et al., 2024; Rogat et al., 2022; Xu

Table 1 Selection of self-assessment scales that assess similar aspects, dimensions or constructs as the QCGE

Scale	Authors	Participants/target group	Based on (theory/evidence)	Content	Sample items
National Survey of Student Engagement (NSSE)*	Indiana University (NSSE, 2016)	Undergraduate students in higher education	Engagement indicators have been developed using a mix of theory and empirical analysis. Engagement is defined and measured as being essentially a matter of student behaviour, which can be observed (Axelson and Flick, 2010)	87 items and four engagement themes: – Academic challenge – Learning with peers – Experiences with faculty – Campus environment	*Academic challenge:* 'During the current school year, how often have you connected your learning to societal problems or issues?' *Learning with peers:* 'During the current school year, how often have you explained course material to one or more students?' *Experiences with faculty:* 'During the current school year, how often have you talked about career plans with a faculty member?' *Campus environment:* 'How much does your institution emphasize the following: providing support to help students succeed academically?'
Cognitive engagement subscale – part of the Motivation and Strategy Use Survey (MSUS)*	Greene and Miller (1996)	College students	Previous models by Nolen (1988) and Meece et al. (1988). Cognitive engagement is defined as 'students' reported use of metacognitive and self-regulatory strategies rather than help-seeking or effort-avoidance strategies' (Meece et al., 1988, p. 515)	Two subdimensions: – Meaningful cognitive engagement – Shallow cognitive engagement	*Meaningful cognitive engagement:* 'I made a plan for achieving the grade I wanted on this exam' *Shallow cognitive engagement:* 'I tried to write down exactly what my instructor said during lectures'

Table 1 (continued)

Scale	Authors	Participants/target group	Based on (theory/evidence)	Content	Sample items
Pair Programming Mutual Engagement (PPME) Questionnaire	Xu and Correia (2024)	Middle school student dyads engaged in pair programming in a computer science course	Items were adapted from the following studies on student engagement: Sinha et al. (2015); Naibert and Barbera (2022); Bryan-Kinns (2013); Edwards et al. (2010)	20 items and four subscales: – Behavioural engagement – Cognitive engagement – Emotional engagement – Social engagement	*Behavioural engagement:* 'I contributed to our project' *Cognitive engagement:* 'we set specific goals for completing the project' *Emotional engagement:* 'I looked forward to today's activity' *Social engagement:* 'I feel my opinions have been taken into account'
Collaborative Learning Engagement Scale	Xu et al. (2024)	Chinese university students in collaborative learning activities	Items were adapted from well-established measures of student engagement, including Fredricks et al. (2004), Gunuc and Kuzu (2015), Pekrun et al. (2002), Muukkonen et al. (2020)	29 items within a three-factor model of collaborative learning engagement: – Behavioural engagement – Cognitive engagement – Emotional engagement	*Behavioural engagement:* 'I actively interact with peers in my group' *Cognitive engagement:* 'I make enough effort to do our group work' *Emotional engagement:* 'When I think about our group work, I feel worried/relaxed'

Table 1 (continued)

Scale	Authors	Participants/target group	Based on (theory/evidence)	Content	Sample items
Group regulation questionnaire	Lai (2021)	Higher education students in project-based learning	The questionnaire was modified from the measurement for self-regulated learning developed by Papamitsiou and Economides (2019)	12 items and four sub-dimensions: – Effort regulation – Goal expectancy – Help-seeking – Time management	*Effort regulation:* 'In this lesson, we try to do the tasks given to us by the teacher' *Goal expectancy:* 'We think that we have paid enough attention to this lesson' *Help-seeking:* 'If we have problems with the project, we ask the trainer for help' *Time management:* 'We try to plan at least once a week to work together on the project'
Scale to assess self-, co-, and socially shared regulation of intrinsic motivation for collaborative activities	Ito and Umemoto (2022)	University students and working adults engaged in collaborative activities	The scale is based on existing scales that assess intrinsic motivational regulation strategies (Ito & Shinto, 2003) and intrinsic motivational regulation strategies of cooperative learning (Umemoto et al., 2018)	Nine items and three subscales: – Self-regulation, – Co-regulation – Social regulation of intrinsic motivation	*Self-regulation:* 'I try to increase my motivation by making the contents of group activities as enjoyable as possible' *Co-regulation:* 'I try to support the motivation of the group member by making the contents of group activities as enjoyable as possible' *Socially shared regulation:* 'I try to support the motivation of the entire group by devising ways to make them find the contents of group activities interesting'

Table 1 (continued)

Scale	Authors	Participants/target group	Based on (theory/evidence)	Content	Sample items
Group metacognition scale	Biasutti and Frate (2018)	Higher education students who participated in an online group study	The scale was developed on the basis of definitions of the dimensions of group metacognition by Schraw and Moshman (1995) and on adaptions of items from existing scales including Garrison and Akyol (2013) and Biasutti (2011)	20 items and four dimensions: – Knowledge of cognition – Planning – Monitoring – Evaluation	*Knowledge of cognition:* 'We know our strengths as learners' *Monitoring:* 'We ask questions to check our understanding' *Evaluation:* 'We make judgments about workload' *Planning:* 'We select the appropriate tools'
Social Metacognition Inventory	Zheng et al., (2023a, 2023b)	Online learning groups of undergraduate students	Operational definitions of social metacognition (Efklides, 2008) with reference to previous scales on individual metacognition (O'Neil & Abedi, 1996), group cohesiveness (Huang, 2009), and metacognition in communities of inquiry (Garrison & Akyol, 2013)	24 items and five dimensions: – Beliefs about others – Awareness of others' thinking – Judgment of others' emotions – Co-regulation of others' thinking – Evaluation of others' thinking	*Beliefs of other persons:* 'I believe that team members can take their own responsibilities for the group work' *Awareness of other persons' thinking:* 'I can follow other members' responses' *Judgment of other persons' emotions:* 'I can judge other persons' emotions by reading the emoticons on QQ' *Co-regulation of each other's thinking:* 'I challenge myself or team members for better solutions' *Evaluation of other persons' thinking:* 'I can judge whether or not the claims proposed by others are correct'

Table 1 (continued)

Scale	Authors	Participants/target group	Based on (theory/evidence)	Content	Sample items
The Learners' Perceived Readiness for Computer-Supported Collaborative Learning questionnaire	Xiong et al. (2015)	University students in CSCL	The scale was developed on the basis of prior research on CSCL and readiness issues and on adaptions from items from existing scales including Schoor and Bannert (2011) and Hung et al. (2010)	39 items and three dimensions: – Motivation for collaborative learning – Prospective behaviors for collaborative learning – Online learning aptitude	*Motivation for collaborative learning:* 'The possible reason I would like to participate in collaborative learning is, because I believe I can do well in the group work' *Prospective behaviors for collaborative learning:* 'If I am doing group work, I would be open to new ideas' *Online learning aptitude:* 'I am able to learn new technologies quickly'
Scale for Student Engagement in Group Projects	Bojanova (2013)	University students in group projects	The scale was developed on the basis of student engagement definitions (e.g. NSSE, 2016) and a focus group method	34 items and four dimensions: – Effort – Teamwork – Motivation – Organisation	'To what extent do the following behaviors, thought, and feelings described you in this group project?' *Effort:* 'Communicated clearly and effectively' *Teamwork:* 'Had fun during team activities' *Motivation:* 'Was inspired to learn and contribute' *Organisation:* 'Rehearsed for project presentation'

Table 1 (continued)

Scale	Authors	Participants/target group	Based on (theory/evidence)	Content	Sample items
Questionnaire on Teamwork (FAT)**	Kauffeld and Frieling (2006)	Working teams such as project teams, interdisciplinary and semi-autonomous working teams	The questionnaire was developed on the basis of two models of team development: Beckhard's (1972) hierarchical model of 'goals', 'roles', 'procedures' and 'relationships', and West's (2012) two fundamental dimensions of team functioning: 'task reflexivity' and 'social reflexivity'	Two scales and respective subscales: – Social reflexivity with the subscales: cohesion and assumption of responsibility – Task reflexivity with the subscales: goal orientation and task mastery	*Cohesion:* 'The team is in the centre and not the individual' *Assumption of responsibility:* 'Everyone contributes equally to the team' *Goal orientation:* 'Our goals are realistic and achievable' *Task mastery:* 'Our priorities are clear'
Team Work Engagement Questionnaire (TWE)**	Costa et al. (2014)	Working teams	The authors derived the TWE scale from the Utrecht Work Engagement Scale (UWES, Schaufeli et al., 2006) which measures individual work engagement, and linguistically adapted and validated it using the first-person plural ('we')	The scale consists of nine items and three dimensions – Vigour – Dedication – Absorption	*Vigour:* 'At our work, we feel bursting with energy' *Dedication:* 'We are proud of the work that we do' *Absorption:* 'We get carried away when we are working'

*Self-assessment scales for individual student engagement (i.e. not in a group context). **Self-assessment scale for engagement (or similar construct) in work teams (i.e. not in educational context)

& Correia, 2024), as well as the above mentioned methodological and practical relevance, we aim to develop a self-assessment scale that addresses the conceptualization of QCGE in a CSCL environment and can measure higher education students' self-assessment of group engagement. To this end, we collected new original data and conducted several methodological steps for validating a new self-assessment scale (QCGE-SAS), which we developed as described in the following sections.

Methods

QCGE-SAS scale development

This study builds on a previous step in the development of a collaborative engagement scale (Paneth et al., 2023). In the previous step, the QCGE self-assessment items were derived from the QCGE rating scheme (Table 2) originally developed by Sinha et al. (2015). Sinha et al., (2015, p. 282) used observable indicators to rate the quality of collaborative group engagement, such as on-/off-task behavior or usage of tools. Thus, we created 3 items for each QCGE dimension (Fig. 1), resulting in a scale of 12 items. For each dimension observable indicators, as displayed in Table 2, were carefully derived from the corresponding definitions and rating criteria.

For behavioral engagement (BE), items were crafted to capture the balance between on-task and off-task behaviors, emphasizing the importance of sustained focus and minimal distractions within the group. For example, items were designed to reflect group members' ability to collectively maintain task-oriented behavior throughout the activity. For social engagement (SE), items were developed to assess inclusion, respectfulness and collaboration. These items highlight the importance of equitable participation and respectful interactions, ensuring that all group members contribute meaningfully and that collaboration is fostered. For cognitive engagement (CE), items focused on the processes of planning, structuring, and monitoring group tasks. Drawing from the rating instructions, these items were framed to evaluate how well the group organises its efforts, sets goals, and monitors progress, including time management and adherence to task structures. Finally, items in the conceptual-to-consequential (CC) dimension were developed to capture the integration of evidence, shared knowledge, and task alignment. These items emphasize the depth and focus of group discussions, assessing whether the group effectively uses shared knowledge and evidence to advance task objectives.

By adapting the language and structure of these items to reflect group-level dynamics, the scale was tailored to comprehensively evaluate collaborative engagement across the four dimensions, providing a robust framework for assessing the quality of group collaboration. These items were then discussed in an expert panel and slightly adjusted in terms of criteria such as comprehensiveness, wording and fit with the original QCGE rating scheme. For example, according to relevant linguistic theory (cf. Pennebaker et al., 2003), the use of words such as 'we' or 'our' makes individuals identify with the group as a unit rather than with themselves as individuals. We therefore adjusted the wording accordingly. An initial validation was then carried out, in which we tested the scale on a sample of student groups. As the results were not satisfactory (i.e., unacceptable reliability), the scale was modified and extended. More specifically, the scale was modified by adding an additional three to four items per dimension, including negatively worded items. These new items were again derived from the original QCGE rating scheme from Sinha et al., (2015, p. 282) and were

Table 2 Definition and rating instructions for QCGE dimensions (adapted from Sinha et al., 2015)

Dimension	Content	Rating instruction
BE	On-task versus off-task behavior	Group members should stay focused on the task, with minimal distractions or off-task behavior among individuals. More off-task behavior decreases the quality of engagement
SE	Inclusion, respectful interaction, collaboration	Contributions should be balanced among members, with respectful and collaborative interactions. Unequal participation or disrespect negatively impacts social engagement
CE	Planning, structuring, task monitoring	The group should develop a clear and structured plan to approach the task, accompanied by effective monitoring (e.g. time management). A lack of planning or monitoring diminishes cognitive engagement
CC	Evidence use, shared knowledge, task focus	Discussions should integrate evidence and shared knowledge to align with task goals. Superficial discussions or inconsistent use of evidence reduce engagement in this dimension

BE, behavioral engagement; *SE*, social engagement; *CE*, cognitive engagement; *CC*, conceptual-to-consequential engagement.

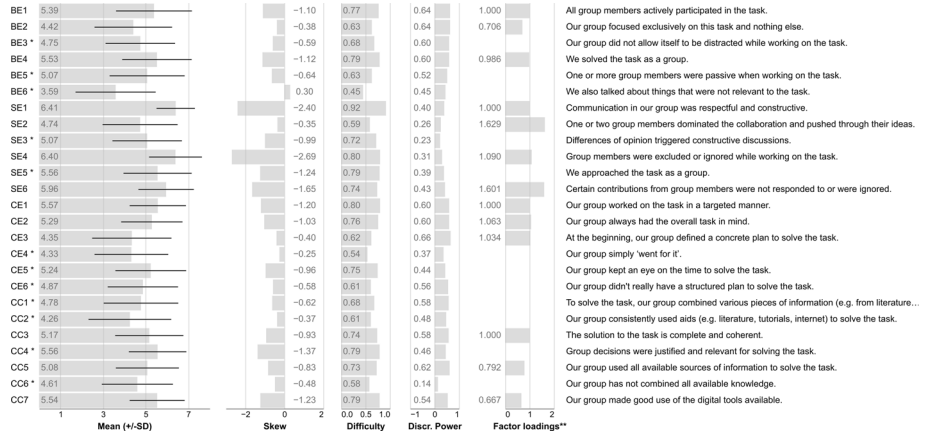

Fig. 1 Mean, standard deviation (SD, horizontal line in leftmost column of bars), skew, difficulty, discriminatory power, and factor loadings are visualised as bars (labelled with the exact values) for each questionnaire item (left: abbreviation; right: full text) in columns. * Items marked with a single asterisk were excluded from the final scale. ** The calculations for the factor loading analysis only included the final selection of items. Therefore, there are only factor loadings for the remaining items

again reviewed in an expert panel. With the additional items, we wanted to ensure that all observable indicators within the four dimensions were included in the scale. For example, the assessment criteria for SE consists of the indicator that group participation is balanced and not dominated by one part of the group. So, we developed an item 'SE2' (Fig. 1) to ask about the subjective experience of the group's collaborative participation during the task. Following this extension and re-testing of the scale, we conducted a second comprehensive round of validation, as described below.

Participants

The sample ($N=246$; Table 3) of this study consisted of undergraduate and postgraduate students from the School of Applied Psychology at the University of Applied Sciences and Arts Northwestern Switzerland. Students were selected to complete the questionnaire as part of the course in which they participated – thus, all students in the sample were part of a learning or study group during a course. In addition, this sample consists of a subsample ($N=87$) of students who participated in a laboratory study using an authentic CSCL task

Table 3 Sample description

	$N=246$
Age (years)	
Mean (SD)	25.22 (5.64)
Range	18–47
Gender	
Male	55
Female	177
Other	5

design. Students worked on a collaborative learning task in groups of three and were asked to complete the questionnaire at the end of their participation in the study (for more details on this specific study design, see Paneth et al., 2024). We included this subsample in order to increase the sample size of this study, and because the task was comparable to the tasks students completed as part of the course work. Students who indicated that they had already completed the scale in a previous data collection (i.e. previous step of validation, see above) were excluded from the sample. Furthermore, only students who completed the entire questionnaire are included in the sample.

Procedure

Participation in the study was voluntary and there were no consequences for students in terms of their course performance. All participants who completed the questionnaire were first asked to complete the informed consent form (approved by the university ethics committee). They were then instructed to answer the questions based on their experiences in their respective study group as part of the course. For the subsample of students from the laboratory study, the procedure was the same: they were asked to complete the questionnaire at the end of their participation in the study to self-assess their perceptions of group work (see detailed study procedure in Paneth et al., 2024).

Measures

Several measures were used to validate the QCGE-SAS, as described below. In addition to the newly developed QCGE-SAS, we used selected and pre-validated scales to test construct validity, which, based on an extensive literature review (Table 1), we found to be particularly similar to specific dimensions of QCGE. However, for the CC subscale of the QCGE-SAS, we did not find a scale or questionnaire measuring a similar construct, as this dimension was newly introduced by Sinha et al. (2015). Nevertheless, to explore some degree of similarity, we compared the subscale with the three other selected scales. In addition, to measure criterion validity, we added an item asking about subjective learning outcomes. All measures are described in detail below.

QCGE Self-Assessment Scale (QCGE-SAS)

The QCGE-SAS, newly developed in this study, was originally worded in German according to the spoken and written language used at the university where the study was conducted. The scale was introduced with the question, 'how would you rate the collaboration in your study group?' This question was accompanied by a help text providing a broader context for the scale: here, participants were instructed to indicate what was most likely to apply to them. The 'learning group' mentioned in the introductory question referred to the group constellation in which they had completed a collaborative task in the course they had attended, and the 'task solution' mentioned in some items referred to the final product (presentation, report, etc.) created at the end of each student group's collaborative task. If there was no final product (yet), they should refer to the preliminary product. Participants

were asked to answer each item on a seven-point Likert scale, where 1 was 'strongly disagree' and 7 was 'strongly agree'. Items were presented in a randomised order.

Teamwork Engagement Scale (TWE)

The Teamwork Engagement scale (TWE; Costa et al., 2014) was selected because of its similarity to the BE subscale of the QCGE-SAS. The TWE was originally worded in English and we translated it into German using an established translation service (DeepL Translate; pro version). The subscales of the TWE were introduced with the question 'to what extent did the following statements apply to the work in your learning group?' Participants were asked to respond to each item on a five-point Likert scale, with 1 representing 'not at all' and 5 representing 'to a very great extent'. The items were presented in a randomised order.

Questionnaire on teamwork (FAT)

The Cohesion subscale of the social reflexivity scale of the Questionnaire on Teamwork (FAT; Kauffeld & Frieling, 2006) was chosen because of its similarity to the SE subscale of the QCGE-SAS. The subscale was introduced with the question 'to what extent did the following statements apply to the work in your learning group?' Participants were asked to respond to each item on a five-point Likert scale, with 1 representing 'not at all' and 5 representing 'to a very great extent'. The items were presented in a randomised order.

Motivation and Strategy Use Survey (MSUS)

The meaningful cognitive engagement subscale of the Motivation and Strategy Use Survey (MSUS; Greene & Miller, 1996) was chosen because of its similarity to the CE subscale of the QCGE-SAS. The MSUS was originally worded in English and we translated it into German using an established translation service (DeepL Translate; pro version). The subscale was introduced with the question 'to what extent did the following statements apply to the work in your group?' Participants were asked to respond to each item on a five-point Likert scale, with 1 representing 'not at all' and 5 representing 'to a very great extent'. The items were presented in a randomised order.

Self-reported learning outcome

Participants were asked: 'How would you rate your group work overall? Please indicate which of the following applies most to you'. They were then presented with the item 'I learned a lot from our group work', which they were asked to rate on a seven-point Likert-type scale, where 1 was 'strongly disagree' and 7 was 'strongly agree'.

Objective learning outcome

The subsample of participants from a previous study (Paneth et al., 2024) were graded on the basis of their collaborative group work. To assess this objective learning outcome, a rating grid was developed to evaluate the collaborative group products. The

grid assessed domain-specific outcomes, including achieving common and individual goals, resolving conflicts, and demonstrating creativity in design. A weighted scoring system graded the models on a scale from 1 (poor) to 6 (excellent) in tenths of a grade, consistent with Swiss university standards. Two experts independently evaluated 19 models, achieving good interrater reliability (ICC = 0.825).

Data analysis

The following analyses were performed using version 4.2.1 of R (R Core Team, 2021) in the R Studio environment (RStudio Team, 2023). We used the full dataset for data analysis, but not all participants completed the additional scales to the QCGE-SAS (see Measures section), so there was missing data for these additional measures. We therefore used all available data for each step of the data analysis. The sample sizes for each step of the analysis are listed with the results.

Item analysis and selection

To validate the QCGE-SAS through item analysis, we adhered to the guidelines of Moosbrugger and Kelava (2007), examining the difficulty, variance, skewness, and discriminative power of each item. We began by analysing all items across the entire scale (Fig. 1) and subsequently omitted items on the basis of the analysis results and further validation processes. Specifically, we conducted an initial item analysis, reliability analysis, and confirmatory factor analysis (CFA) for all items. We then iteratively removed items that demonstrated: (1) suboptimal performance in the item analysis (see 'Item Analysis' in the Results section); (2) low reliability (see 'Reliability' in the Results section); or (3) poor factor loadings and model fit in the CFA (see 'Confirmatory Factor Analysis' in the Results section). The final selection of items, as presented in Fig. 1, was used for the concluding reliability analysis and CFA as described below.

Reliability

To test reliability, we calculated the internal consistency of each item using Cronbach's alpha. However, as the use of Cronbach's alpha to obtain model reliability has proven controversial (Tavakol & Dennick, 2011), we additionally calculated the composite reliability (Tentama & Anindita, 2020) of each QCGE-SAS dimension based on the structural equation model from the CFA (see next section).

Confirmatory Factor Analysis (CFA)

We conducted a CFA to test whether the four-factor structure proposed by Sinha et al. (2015) was a good fit. To do this, we ran the full model consisting of all the items shown in Fig. 1. To fit the proposed four-factor structure, each item would be the observed variable loading on the corresponding dimension of the QCGE-SAS.

For the CFA we used the R package lavaan (Rosseel et al., 2017). Since the scale of the QCGE-SAS indicates a seven-point Likert scale and can therefore be viewed as a continuous scale, we chose maximum likelihood as the estimator. Before running the CFA, we tested the multivariate normality of the items (Mardia et al., 2024). The fit of the model

was evaluated using several fit indices, including the Chi-squared test of model fit, the root mean square error of approximation (RMSEA), the standardized root mean square residual (SRMR), the comparative fit index (CFI), and the Tucker–Lewis index (TLI).

Construct validity

To test the construct validity of the QCGE-SAS subscales (i.e. BE, SE, CE, and CC), we calculated various correlations with the additional pre-validated scales that we had added to the questionnaire (see Measures section). These scales were specifically chosen because they align conceptually with our constructs, enabling us to assess convergent validity. We compared the results of the BE subscale with the results of the TWE subscales vigour, dedication, and absorption (Costa et al., 2014); the results of the SE subscale with the results of the cohesion subscale of the FAT (Kauffeld & Frieling, 2006); and the results of the CE subscale with the results of the meaningful cognitive engagement subscale of the Motivation and Strategy Use Survey (Greene & Miller, 1996). For CC, where no directly comparable scale was identified, we explored correlations with all additional scales mentioned above to gain insights into its relationships across constructs. To prepare for correlational construct validity analysis, we first calculated mean scores for all four QCGE-SAS subscales and analysed the distribution of scores to ensure adherence to normality assumptions. Finally, we calculated correlations among the QCGE-SAS subscales to test whether the four dimensions of the QCGE construct were interrelated, providing further evidence of the coherence and distinctiveness of these dimensions.

Criterion validity

In terms of criterion validity, we assumed that a higher QCGE would be associated with better learning outcomes (Liu et al., 2022; Sinha et al., 2015; Xing et al., 2022). Therefore, we calculated correlation coefficients between the four QCGE-SAS subscales and self-reported learning outcomes (see Measures section). We also conducted a subsample analysis including the participants that were graded on their collaborative group work (see Measures section). To analyse the relationship between the QCGE-SAS subscales and the objective learning outcome, we computed a linear regression model. The model included the four dimensions of engagement measured in the questionnaire as independent variables and the group task grade as the dependent variable. This approach allowed us to assess the unique contribution of each engagement dimension to the grade while controlling for the others.

Results

The results of the QCGE-SAS validation steps are reported below. Most of the results are visualised in Figs. 1 and 2, and we refer to these figures at the relevant points in the Results section.

Item analysis

The results of the item analysis are shown in Fig. 1. The skewness of the items indicates that the participants overall tended to report high levels of engagement. Similarly, the item

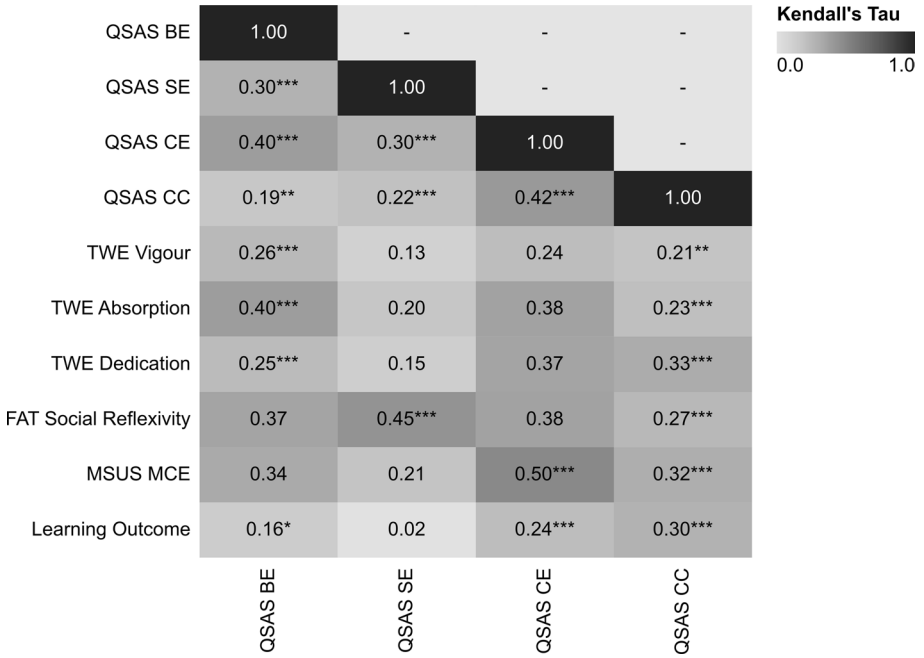

Fig. 2 Kendall's tau between and within the QCGE-SAS and the TWE subscales, the FAT social reflexivity/cohesion subscale, the meaningful cognitive engagement subscale (MSUS MCE) and the self-reported learning outcome. ***$=p<0.001$; **$=p<0.01$; *$=p<0.05$. We calculated p-values only for the tests regarding the construct and criterion validity analysis and between the QCGE-SAS subscales

difficulty tends to be high for most of the items, indicating that participants report high levels of engagement. In terms of discriminatory power, the items show a range of low to high values – for example, items from the SE subscale show rather low discriminatory power (e.g. SE3); whereas, the values for the CE subscale are rather high (e.g. CE3).

Reliability

The results of the reliability analysis are as follows: the Cronbach's alpha coefficients indicate mostly acceptable coefficients, namely for BE $\alpha=0.754$; SE $\alpha=0.560$; CE $\alpha=0.775$; and CC $\alpha=0.738$. That is, BE, CE, and CC indicate acceptable internal consistency, while the coefficient for the SE subscale does not exceed the commonly accepted threshold of 0.70, indicating unacceptable internal consistency (Nunnally & Bernstein, 1994). The composite reliability results obtained from the CFA show reliability scores for BE = 0.76, SE = 0.61, CE = 0.80 and CC = 0.72, confirming the internal consistency results that the BE, CE, and CC subscales have good reliability as they exceed the threshold of 0.70 (Tentama & Anindita, 2020), while the SE subscale again shows a lower reliability score.

In other words, the reliability analysis shows that the BE, CE, and CC subscales have acceptable reliability in terms of both internal consistency and composite reliability, supporting their validity in measuring the intended construct. However, the SE subscale has an unacceptable reliability coefficient, indicating that this subscale may not consistently measure the construct.

Confirmatory factor analysis

The results of the Mardia's skewness test indicate a significant deviation from normality, as indicated by the skewness statistic of 5502.4 with $p < 0.001$. This indicates that the data are not symmetrically distributed around the mean. Similarly, the kurtosis test produced a statistic of 22.01 with $p < 0.001$, indicating a significant deviation in the tail behaviour of the distribution compared with a normal distribution. Therefore, we chose to use a robust estimation method with maximum likelihood parameter estimates with standard errors and a mean adjusted chi-squared test statistic (MLM) method that adjusts for non-normality (Satorra & Bentler, 1994).

Three participants were excluded from the analysis owing to missing responses, resulting in a sample size of $N = 243$. The factor loadings of the final item are presented in Fig. 1. The fit of the model was assessed using several goodness of fit indices. The $\chi 2$ test showed 140.51 (df = 48), with $p = 0.001$. As the $\chi 2$ test is sensitive to sample size, additional fit indices were considered. The robust comparative fit index (CFI) was CFI = 0.933, indicating good fit, and the robust Tucker–Lewis index (TLI) was TLI = 0.911, also indicating good fit (Hu & Bentler, 1999). The robust root mean square error of approximation (RMSEA) was RMSEA = 0.069 with a 90% confidence interval of 0.051–0.088, indicating an acceptable fit (Hu & Bentler, 1999). Also, the standardised root mean square residual (SRMR) was SRMR = 0.062, which is below the recommended threshold of 0.080 (Hu & Bentler, 1999), indicating a good fit. In summary, the CFA shows that the four-factor structure of the QCGE-SAS fits well.

Construct validity

Construct validity analysis was conducted on a subsample ($N = 112$, see Data Analysis section). As all QCGE-SAS subscales showed a non-normal distribution with the Shapiro–Wilk normality test ($p < 0.001$), we computed the non-parametric Kendall's tau test for all four comparisons. The results, as shown in Fig. 2, are reported below.

Regarding the correlations between the four subscales of the QCGE-SAS (i.e. BE, SE, CE, and CC), the results indicate a positive, statistically significant relationship. The results of the Kendall rank correlation test for the correlations between BE and the Team Work Engagement subscales of absorption, vigour, and dedication show positive, statistically significant relationships. Between SE and the subscale cohesion/social reflexivity (FAT), the results indicate a positive, statistically significant relationship, and between CE and the subscale meaningful cognitive engagement, the results indicate a positive, statistically significant relationship. Finally, the results for the correlations between CC and all of the above scales show positive, statistically significant relationships.

In summary, the observed correlations between all four subscales of the QCGE-SAS with the reported scales measuring similar constructs provide strong support for the convergent construct validity of the QCGE-SAS subscales used in this analysis. Furthermore, the correlations within the QCGE-SAS between the four dimensions (i.e. BE, SE, CE, and CC) suggest that the subscales measure similar but distinct constructs.

Criterion validity

Self-reported learning outcome

Criterion validity regarding the self-reported learning outcome analysis was conducted on a subsample ($N = 116$, see Data analysis section). Given the non-normal distribution of the

QCGE-SAS subscales, we again calculated the non-parametric Kendall's tau test for all four comparisons. The results, as shown in Fig. 2, indicate a statistically significant relationship between BE and self-reported learning outcomes, and no statistically significant relationship between SE and self-reported learning outcomes. The results for the correlations between CE and CC with self-reported learning outcomes, on the other hand, show positive, statistically significant relationships.

In other words, the criterion validity analysis revealed different levels of correlation between the QCGE-SAS subscales and the self-reported learning outcomes, and the relationships for CE and CC are stronger than the relationships for BE and SE.

Objective learning outcome

Regarding the objective learning outcome, criterion validity was conducted on another subsample ($N=81$, see Data analysis section). The results indicate (Table 4) that the model explains a statistically significant proportion of variance (R2 = 0.12, F(4, 76) 2.71, $p=0.036$, adj. R2 = 0.08). Notably, the predictor for CC demonstrates a significant positive effect (Table 4), suggesting its unique and meaningful contribution to the model. Other predictors (BE, SE, and CE) did not show statistically significant effects.

Discussion

In this study, we asked how the quality of collaborative group engagement (QCGE) can be measured using a self-assessment scale with four dimensions: behavioural, social, cognitive, and conceptual-to-consequential engagement in higher education contexts. To answer this research question, we successfully developed and validated the 13-item QCGE-SAS (Fig. 1). The results reveal several key insights, which we discuss below.

While item reliability showed good internal consistency for the BE, CE, and CC subscales, this was not the case for the SE subscale. Consistent with these findings, composite reliability indicated acceptable reliability for all but the SE subscale. In line with our previous research (Paneth et al., 2023, 2024), the social dimension of QCGE may be the most challenging dimension to assess through self-assessment, but also through other measures such as observations owing to social desirability bias (Nederhof, 1985) and motivational factors in study and laboratory settings (Van Lange et al., 2011). For example, in our previous work, we reported low variance for observed (i.e. rated) SE (Paneth et al., 2023, 2024) as well as discrepancies between self-assessment and observed SE (Paneth et al., 2023).

Table 4 Linear regression results for the criterion validity of the objective learning outcome ($N=81$)

Predictor	B	95% CI for B	t (df = 76)	p-Value
BE	−0.03	[−0.27, 0.21]	−0.27	0.79
SE	−0.06	[−0.29, 0.16]	−0.57	0.57
CE	0.04	[−0.25, 0.33]	0.27	0.788
CC	0.32	[0.05, 0.59]	2.33	0.022*

B, unstandardized coefficient; *CI*, confidence interval; *BE*, behavioral engagement; *SE*, social engagement; *CE*, cognitive engagement; *CC*, conceptual-to-consequential engagement. * $p<0.05$.

In terms of reliability and validity of the QCGE-SAS, the CFA shows good fit indices, indicating that the overall model structure fits the data well. This suggests that although the SE subscale shows lower reliability, the four-factor structure proposed by Sinha et al. (2015) is sound. Moreover, correlation analysis between the four dimensions of the QCGE-SAS suggests that four similar but distinct constructs are being measured. This leads to a final selection of items, based on which we present the proposed form of the QCGE-SAS. Interestingly, the mostly negatively worded items of the SE subscale contributed more to the model fit than the positively worded items, suggesting that they may better discriminate between social engagement of groups. These findings could again be explained by a certain social desirability bias (Nederhof, 1985), which suggests that social desirability is less likely to be triggered when items are worded negatively than when they are worded positively. For example, in SE1 ('the communication in our group was respectful and constructive'), participants in one study group might agree and, in the meantime, experience dominant behaviour from one part of the group and report this in the negatively worded SE2 ('one or two group members dominated the collaboration and pushed through their ideas'). This may have resulted in a more nuanced indication of SE than if the items were all positively worded.

Construct validity was supported by significant correlations with the related scales. Criterion validity results indicated that learning outcomes correlated more strongly with CC than with BE, SE, and CE (i.e. not significant for the latter three). These findings are consistent with Sinha et al. (2015), suggesting that while high levels of SE and BE are essential prerequisites for the formation of high levels of CE and CC, they alone are not sufficient to ensure overall engagement and thus, learning success. Learning groups may be on-task (i.e. BE) and socially engaged (i.e. SE) without pursuing an action plan (i.e. CE) and being consequentially engaged (i.e. CC). Overall, these findings suggest that the CC subscale may be a particularly useful indicator of perceived learning effectiveness, while the BE, SE, and CE subscales may require further investigation or refinement to better capture relevant aspects of the self-reported learning outcome.

In summary, the good fit indices from the factor analysis support the overall construct validity of the scale, suggesting that it measures the intended construct well. The results of the item, reliability and factor analyses suggest that while the SE subscale may need further refinement to improve internal consistency, the BE, CE, and CC subscales are reliable and valid.

The dimensionality of collaborative engagement

Our results are consistent with the four-factor structure proposed by Sinha et al. (2015), which is an encouraging step towards capturing the multidimensional nature of QCGE as a construct. However, alternative frameworks and nomenclatures provide valuable perspectives on the dimensionality of collaborative engagement. Rogat et al. (2022), for example, proposed a five-dimensional model comprising behavioral, collaborative, socio-emotional, metacognitive and disciplinary engagement. This framework expands on Sinha et al. (2015) by redefining cognitive as metacognitive and social as socio-emotional engagement, as well as introducing disciplinary engagement. These dimensions emphasize, respectively, the regulation of collaborative processes and the integration of domain-specific practices and knowledge. These additions reflect the complex interplay between individual and group-level factors in collaborative learning contexts.

The disciplinary dimension underscores the importance of grounding collaborative efforts within the specific epistemological and practical demands of a domain. Meanwhile, the collaborative and socio-emotional dimensions emphasize the necessity of fostering group cohesion and maintaining positive interpersonal dynamics, which are critical for sustaining long-term engagement.

These distinctions suggest that while the QCGE framework effectively captures broad dimensions of collaborative engagement, further refinement may enhance its applicability and alignment with domain-specific and process-oriented learning environments. Incorporating these dimensions into future iterations of the QCGE-SAS could provide a more comprehensive understanding of engagement, particularly in contexts where metacognitive and disciplinary processes are central to group success.

The confirmation of the QCGE-SAS's four-factor structure, considering these alternative perspectives, demonstrates its robustness while also highlighting areas for potential development. Future research should consider integrating metacognitive and disciplinary dimensions or exploring how existing dimensions interact dynamically to influence group outcomes. For example, disciplinary engagement may serve as a mediator between cognitive and behavioral engagement, channelling effort and persistence into domain-specific achievements.

By capturing individual perceptions across behavioral, social, cognitive, and conceptual-to-consequential dimensions, the QCGE-SAS offers a practical and scalable tool for assessing collaborative engagement. As engagement frameworks continue to evolve, the QCGE-SAS can provide a foundation for further refinements, ensuring its relevance and utility across diverse educational contexts. Future studies should aim to validate additional dimensions and investigate their role in fostering high-quality collaboration.

Implications

With this new self-assessment instrument, we contribute to a more precise operationalisation of the four dimensions of group engagement and to the set of valid and reliable QCGE measures. Compared with existing scales that measure similar constructs or dimensions to QCGE, such as the Pair Programming Mutual Engagement (PPME) questionnaire (Xu & Correia, 2024) or the Collaborative Learning Engagement Scale (Xu et al., 2024) (Table 1), the QCGE-SAS offers distinct differences. It explicitly measures the QCGE construct across its four dimensions (i.e. BE, SE, CE, and CC) while being specifically designed for a CSCL setting. Unlike the PPME questionnaire, which is tailored to middle school contexts and focuses on dyadic interactions, the QCGE-SAS is not limited to two-member group structures and instead takes the entire group as the unit of analysis. Similarly, while other scales such as the PPME questionnaire and the Collaborative Learning Engagement Scale address engagement, they are not explicitly grounded in the theoretical underpinnings of CSCL. Moreover, measuring the dimensions of QCGE from the learner's perspective allows for detailed future research and theory building, and can deepen scientific knowledge about different perceptions of QCGE. For example, we could compare observational perspectives and subjective or intersubjective (i.e. group) perspectives, both of which provide unique views of QCGE, but taken together provide a more comprehensive picture of the construct than either perspective alone (Paneth et al., 2023; Pekrun, 2020; Vriesema & McCaslin, 2020), thus contributing to comprehensive multi-method approaches. In addition, within-group perspectives of group members could be explored in terms of whether group members'

perspectives are homogeneous or heterogeneous, and what consequences this has for communication and collaborative learning. The QCGE-SAS contributes to this further research, which also informs CSCL practice.

In terms of practical implications for CSCL learning groups, self-assessment questionnaires can be used in groups not only to identify existing processes in learning groups but also as a basis for adjustments and interventions in group work. Previous research has shown that adjustments of interpersonal processes within groups are more effective when group members share a *common ground* (Medina & Stahl, 2021) or *shared representations* about their processes, enabling them to perform better (e.g. Ainsworth & Chounta, 2021). However, the use of self-assessment questionnaires in learning groups does not automatically lead to positive effects on group processes. Rather, such positive effects are dependent on the design and adaptation of group-related processes, which requires an assessment of the situation in the first place (Burke et al., 2006). Our self-assessment described here offers the possibility of identifying group-related indicators of engagement and reporting them back to the group – a central basis for adapting engagement-related group processes and improving learning processes. Learning environments often lack both process and socio-emotional feedback and the QCGE-SAS can be used as a feedback tool in CSCL practice, for example as part of automated group engagement feedback tools for student groups (Zheng et al., 2023a, 2023b) or as a component of group awareness tools (Chen et al., 2024; Schnaubert & Bodemer, 2022), providing student groups with information about their group learning process and thus improving learning outcomes (e.g. Zheng et al., 2023a, 2023b).

Previous related research (Buder, 2011) has suggested group awareness tools with a focus on displaying (e.g. feedback) and monitoring (e.g. regulating awareness). More specifically, Isohätälä et al. (2017) used tablet tools for targeted socio-emotional regulation in student groups and investigated their effects on collaborative learning. Students were asked to work collaboratively on complex mathematics tasks and to use the tool to reflect on their cognitive capabilities, motivation, and emotions concerning the task. Video analyses from this study reveal how student participation and responsiveness fluctuated during the collaborative learning process. Nguyen et al. (2025) investigated the effectiveness of socio-emotional learning practices in a gamified flipped classroom setting (which combined video lessons with earning experience points). Students of an experimental group were instructed to be self-aware, express emotions, provide positive feedback, and provide 'compliment before feedback' or to share at least '…one positive reflection before leaving class' etc. (p. 7). The controls studied the same contents within the same setting, but without such group-oriented regulation. The authors report positive effects on learning achievement and communication from their mixed methods study based on questionnaire data and interview analyses. Such examples show what the likely interactions between collaborative group engagement and different types of mediating technology could be. The QCGE-SAS could be used as a regulation tool, covering the four different dimensions of QCGE systematically. Future research could demonstrate the effects of such tool use in technology and AI-enhanced collaborative learning settings.

In university teaching, the QCGE-SAS can also be integrated into formative group assessment, for example as a basis for assessing group work as a 'process grade'. As group performance in academic settings is often less valued and incentivised compared with individual performance (Shimizu et al., 2020), integrating group assessment alongside individual assessment reinforces group work in CSCL, and thus supports future skills (Forsell et al., 2020; Tumpa et al., 2022).

Limitations

This study has several limitations that should be considered when interpreting the results. First, the sample size is relatively small. On the one hand, according to Bentler and Chou (1987), the recommended sample size for structural equation modelling is 5–10 subjects per model parameter; in our final model we use 45 model parameters and 243 subjects, so our sample size meets this criterion. On the other hand, more recent guidelines recommend larger sample sizes to ensure robustness and generalisability of results, suggesting that our sample may be at the lower end of the acceptable range. For example, for non-normally distributed data and the use of maximum likelihood with robust standard errors and the Satorra–Bentler scaled chi-squared correction (MLM) estimation method, a sample size greater than $N > 250$ would be optimal (Hu & Bentler, 1999). In addition, the results of our final model based on the selected items should be confirmed in a new sample. Therefore, future research should confirm our findings with a larger sample.

Second, the reliability of the social engagement (SE) subscale is suboptimal. Although efforts have been made to address these reliability issues, the reliability could still be improved. This may be owing to the complexity of the SE QCGE dimension, which consists of different criteria such as cohesion, respectful conversation, and inclusion, as well as the influence of a potential social desirability bias, as discussed above. Future research could, for example, explore the possibility of splitting the SE dimension or consider alternative or additional items to increase the reliability of the SE dimension used in this study.

Third, the assessment of individual self-reflection on group processes poses a significant challenge. Stahl and Hakkarainen (2021) discuss the distinction between subjective, intersubjective, and inter-objective perspectives on CSCL. Our scale is likely to fall between subjective and intersubjective perspectives, raising questions about the validity and feasibility of capturing individual reflections on group dynamics. Future research should explore this issue in more depth to better understand and address the complexities involved in assessing individual self-reflection within group processes, for example, by aggregating data at the group level or using complementary methodological approaches such as group discussions.

Finally, the reliance on self-reported or objective measures of learning outcomes and engagement assessment presents separate challenges. Self-reported measures are inherently subjective and may be influenced by social desirability bias, inaccurate self-perception, or retrospective distortions. For instance, Pike (1996) discusses the limitations of using students' self-reports of academic development as proxies for traditional achievement measures, highlighting potential biases and inaccuracies. These factors can reduce the validity of the reported engagement and learning outcomes. Conversely, objective measures of learning outcomes, while more standardized, may fail to capture the nuanced and context-specific aspects of learning that participants value or experience. Caspersen et al. (2017) emphasize the challenges in measuring learning outcomes, noting that standardized assessments may not fully encompass the diverse skills and knowledge students acquire. In addition, our use of a single-item self-report for measuring the learning outcome introduces another limitation. Single-item measures, while simple and easy to administer, may lack reliability and fail to capture the complexity and multidimensionality of the construct they aim to measure (Diamantopoulos et al., 2012). Future research should confirm our findings by comparing the QCGE-SAS subscales with reliable and valid measurements of learning outcomes. Furthermore, we suggest that revisions to the QCGE-SAS focus on improving the internal consistency of the social engagement subscale.

Conclusion

We successfully developed and validated a self-assessment scale to measure the quality of collaborative group engagement (QCGE-SAS). Our findings confirm the four-factor structure proposed by Sinha et al. (2015) and demonstrate that the QCGE-SAS can successfully capture the multidimensional nature of QCGE. As such, this research contributes to the operationalisation of the four dimensions of QCGE and expands the set of valid and reliable QCGE measures. To this end, QCGE-SAS can complement existing methods and be used within multi-method approaches to provide a comprehensive picture of the complex group processes that are indicative of QCGE. In addition to being a valuable tool for academic research, the QCGE-SAS can have important practical implications for CSCL learning groups. To this end, QCGE-SAS can complement existing methods and be used within multi-method approaches to provide a comprehensive picture of the complex group processes that are indicative of QCGE. In addition to being a valuable tool for academic research, the QCGE-SAS can have important practical implications for CSCL learning groups. Used as a group awareness and feedback tool, or as a tool for formative assessment of group performance, it can shed light on group processes, inform adjustments, and guide interventions.

Acknowledgements We would like to thank all the students who participated in this study and Valentina Vogel for proofreading and her support.

Author contributions Loris T. Jeitziner*: conceptualization, investigation, methodology, writing – review and editing, original draft preparation, data curation, formal analysis and visualization. Lisa Paneth*: conceptualization, investigation, methodology, writing – review and editing, and original draft preparation. Oliver Rack: conceptualization, writing – review and editing, and original draft preparation. Susanne Bleisch: conceptualization, writing – review and editing, and visualization. Carmen Zahn: conceptualization, writing – review and editing, original draft preparation and supervision.

Funding Open access funding provided by FHNW University of Applied Sciences and Arts Northwestern Switzerland. This research was supported by the Swiss National Science Foundation under National Research Program 77 and research project no. 407740_187258 and by the University of Applied Sciences Northwestern Switzerland, School of Applied Psychology and School of Architecture, Civil Engineering, and Geomatics.

Data availability The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Declarations

Conflict of interests The authors declare no competing interests.

Ethics vote The study was approved by the ethics committee of the School of Applied Psychology, University of Applied Sciences and Arts Northwestern Switzerland. Approved application number: EAaFE200622a. Written informed consent was obtained from the participants.

References

Ainsworth, S. E., & Chounta, I.-A. (2021). The roles of representation in computer-supported collaborative learning. In U. Cress, C. Rosé, A. F. Wise, spsampsps J. Oshima (Eds.), *International Handbook of Computer-Supported Collaborative Learning* (Vol. 19). Springer.

Axelson, R. D., & Flick, A. (2010). Defining student engagement. *Change: The Magazine of Higher Learning*, *43*(1), 38–43. https://doi.org/10.1080/00091383.2011.533096

Beckhard, R. (1972). Optimizing team-building efforts. *Journal of Contemporary Business, 1*(3), 23–32.

Bentler, P. M., & Chou, C.-P. (1987). Practical issues in structural modeling. *Sociological Methods & Research, 16*(1), 78–117.

Biasutti, M. (2011). The student experience of a collaborative e-learning university module. *Computers & Education, 57*(3), 1865–1875. https://doi.org/10.1016/j.compedu.2011.04.006

Biasutti, M., & Frate, S. (2018). Group metacognition in online collaborative learning: Validity and reliability of the group metacognition scale (GMS). *Educational Technology Research and Development, 66*(6), 1321–1338. https://doi.org/10.1007/s11423-018-9583-0

Bojanova, I. (2013). Development of a reliable, valid, multi-dimensional measure of student engagement in group projects. *IEEE Frontiers in Education Conference (FIE), 2013*, 239–245. https://doi.org/10.1109/FIE.2013.6684825

Buder, J. (2011). Group awareness tools for learning: Current and future directions. *Computers in Human Behavior, 27*(3), 1114–1117. https://doi.org/10.1016/j.chb.2010.07.012

Bryan-Kinns, N. (2013). Mutual engagement and collocation with shared representations. *International Journal of Human-Computer Studies, 71*(1), 76–90. https://doi.org/10.1016/j.ijhcs.2012.02.004

Burke, C. S., Stagl, K. C., Salas, E., Pierce, L., & Kendall, D. (2006). Understanding team adaptation: A conceptual analysis and model. *Journal of Applied Psychology, 91*(6), 1189.

Caspersen, J., Smeby, J. C., & Olaf Aamodt, P. (2017). Measuring learning outcomes. *European Journal of Education, 52*(1), 20–30. https://doi.org/10.1111/ejed.12205

Chen, D., Zhang, Y., Luo, H., Zhu, Z., Ma, J., & Lin, Y. (2024). Effects of group awareness support in CSCL on students' learning performance: A three-level meta-analysis. *International Journal of Computer-Supported Collaborative Learning, 19*(1), 97–129. https://doi.org/10.1007/s11412-024-09418-3

Costa, P., Margarida Passos, A., & Bakker, A. (2014). Empirical validation of the Team Work Engagement construct. *Journal of Personnel Psychology, 13*(1), 34–45. https://doi.org/10.1027/1866-5888/a000102

Curşeu, P. L., Rusu, A., Maricuţoiu, L. P., Vîrgă, D., & Măgurean, S. (2020). Identified and engaged: A multi-level dynamic model of identification with the group and performance in collaborative learning. *Learning and Individual Differences, 78*, 101838. https://doi.org/10.1016/j.lindif.2020.101838

Diamantopoulos, A., Sarstedt, M., Fuchs, C., et al. (2012). Guidelines for choosing between multi-item and single-item scales for construct measurement: A predictive validity perspective. *Journal of the Academy of Marketing Science, 40*, 434–449. https://doi.org/10.1007/s11747-011-0300-3

Détienne, F., Baker, M., & Burkhardt, J. M. (2012). Perspectives on quality of collaboration in design. *CoDesign, 8*(4), 197–199. https://doi.org/10.1080/15710882.2012.742350

Edwards, R. L., Stewart, J. K., & Ferati, M. (2010). Assessing the effectiveness of distributed pair programming for an online informatics curriculum. *ACM Inroads, 1*(1), 48–54. https://doi.org/10.1145/1721933.1721951

Efklides, A. (2008). Metacognition: Defining its facets and levels of functioning in relation to self-regulation and co-regulation. *European Psychologist, 13*(4), 277–287.

Ferreira, H. A., spsampsps Saraiva, M. (2019). Subjective and objective measures. In H. Ayanoğlu spsampsps E. Duarte (Eds.), *Emotional Design in Human-Robot Interaction* (pp. 143–159). Springer International Publishing. https://doi.org/10.1007/978-3-319-96722-6_9

Forsell, J., Forslund Frykedal, K., & Hammar Chiriac, E. (2020). Group work assessment: Assessing social skills at group level. *Small Group Research, 51*(1), 87–124. https://doi.org/10.1177/1046496419878269

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74*(1), 59–109. https://doi.org/10.3102/00346543074001059

Garrison, D. R., & Akyol, Z. (2013). Toward the development of a metacognition construct for communities of inquiry. *The Internet and Higher Education, 17*, 84–89. https://doi.org/10.1016/j.iheduc.2012.11.005

Greene, B. A., & Miller, R. B. (1996). Influences on achievement: Goals, perceived ability, and cognitive engagement. *Contemporary Educational Psychology, 21*(2), 181–192. https://doi.org/10.1006/ceps.1996.0015

Gunuc, S., & Kuzu, A. (2015). Student engagement scale: Development, reliability and validity. *Assessment & Evaluation in Higher Education, 40*(4), 587–610. https://doi.org/10.1080/02602938.2014.938019

Heilporn, G., Raynault, A., & Frenette, É. (2024). Student engagement in a higher education course: A multidimensional scale for different course modalities. *Social Sciences & Humanities Open, 9*, 100794. https://doi.org/10.1016/j.ssaho.2023.100794

Hmelo-Silver, C. E., spsampsps Jeong, H. (2021). An overview of CSCL methods. In U. Cress, C. Rosé, A. F. Wise, spsampsps J. Oshima (Eds.), *International Handbook of Computer-Supported Collaborative Learning* (Vol. 19). Springer.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55.

Huang, C.-C. (2009). Knowledge sharing and group cohesiveness on performance: An empirical study of technology R&D teams in Taiwan. *Technovation, 29*(11), 786–797. https://doi.org/10.1016/j.technovation.2009.04.003

Hung, M.-L., Chou, C., Chen, C.-H., & Own, Z.-Y. (2010). Learner readiness for online learning: Scale development and student perceptions. *Computers & Education, 55*(3), 1080–1090. https://doi.org/10.1016/j.compedu.2010.05.004

Isohätälä, J., Järvenoja, H., & Järvelä, S. (2017). Socially shared regulation of learning and participation in social interaction in collaborative learning. *International Journal of Educational Research, 81*, 11–24. https://doi.org/10.1016/j.ijer.2016.10.006

Ito, T., & Shinto, T. (2003). Construction of the self-motivational strategies scale for junior high school students. *The Japanese Journal of Psychology, 74*(3), 209–217. https://doi.org/10.4992/jjpsy.74.209

Ito, T., & Umemoto, T. (2022). Self-regulation, co-regulation, and socially shared regulation of motivation for collaborative activity: Comparison between university students and working adults. *Japanese Psychological Research, 64*(4), 397–409. https://doi.org/10.1111/jpr.12337

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

Jeitziner, L. T., Paneth, L., Rack, O., & Zahn, C. (2024a). 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. (2024b). *Automating the Assessment of Collaborative Engagement Using Natural Language Processing*. Center for Open Science. https://doi.org/10.31234/osf.io/5437v

Kauffeld, S., & Frieling, E. (2006). Der Fragebogen zur Arbeit im Team (F-A-T). *Zeitschrift Für Arbeits- und Organisationspsychologie a&o, 45*(1), 26–33. https://doi.org/10.1026/0932-4089.45.1.26

Lai, C. (2021). Effects of the group-regulation promotion approach on students' individual and collaborative learning performance, perceptions of regulation and regulation behaviours in project-based tasks. *British Journal of Educational Technology, 52*(6), 2278–2298. https://doi.org/10.1111/bjet.13138

Liu, S., Liu, S., Liu, Z., Peng, X., & Yang, Z. (2022). Automated detection of emotional and cognitive engagement in MOOC discussions to predict learning achievement. *Computers & Education, 181*, 104461. https://doi.org/10.1016/j.compedu.2022.104461

Mardia, K. V., Kent, J. T., & Taylor, C. C. (2024). *Multivariate analysis* (Vol. 88). John Wiley & Sons.

Medina, R., spsampsps Stahl, G. (2021). Analysis of Group Practices. In U. Cress, C. Rosé, A. F. Wise, spsampsps J. Oshima (Eds.), *International Handbook of Computer-Supported Collaborative Learning* (Vol. 19). Springer.

Meece, J. L., Blumenfeld, P. C., & Hoyle, R. H. (1988). Students' goal orientations and cognitive engagement in classroom activities. *Journal of Educational Psychology, 80*(4), 514–523. https://doi.org/10.1037/0022-0663.80.4.514

Meier, A., Spada, H., & Rummel, N. (2007). A rating scheme for assessing the quality of computer-supported collaboration processes. *Computer Supported Learning, 2*, 63–86. https://doi.org/10.1007/s11412-006-9005-x

Moosbrugger, H., & Kelava, A. (2007). *Testtheorie und fragebogenkonstruktion* (Vol. 537). Springer.

Muukkonen, H., Lakkala, M., Lahti-Nuuttila, P., Ilomäki, L., Karlgren, K., & Toom, A. (2020). Assessing the development of collaborative knowledge work competence: Scales for higher education

course contexts. *Scandinavian Journal of Educational Research, 64*(7), 1071–1089. https://doi.org/10.1080/00313831.2019.1647284

Naibert, N., & Barbera, J. (2022). Development and evaluation of a survey to measure student engagement at the activity level in general chemistry. *Journal of Chemical Education, 99*(3), 1410–1419. https://doi.org/10.1021/acs.jchemed.1c01145

Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology, 15*(3), 263–280. https://doi.org/10.1002/ejsp.2420150303

Nguyen, V. T. T., Chen, H., & Nguyen, V. T. K. (2025). Integrating social–emotional learning into gamified flipped classrooms: Impacts on emotion regulation, achievement and communication tendency. *Journal of Computer Assisted Learning, 41*(1), e13109. https://doi.org/10.1111/jcal.13109

Nolen, S. B. (1988). Reasons for studying: Motivational orientations and study strategies. *Cognition and Instruction, 5*(4), 269–287. https://doi.org/10.1207/s1532690xci0504_2

NSSE. (2016). *About NSSE*. Evidence-based improvement in higher education. https://nsse.indiana.edu/

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed). McGraw-Hill.

O'Neil, H. F., & Abedi, J. (1996). Reliability and validity of a state metacognitive inventory: Potential for alternative assessment. *The Journal of Educational Research, 89*(4), 234–245. https://doi.org/10.1080/00220671.1996.9941208

Paneth, L., Jeitziner, L., 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/10.22318/cscl2023.134087

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

Papamitsiou, Z., & Economides, A. A. (2019). Exploring autonomous learning capacity from a self-regulated learning perspective using learning analytics. *British Journal of Educational Technology, 50*(6), 3138–3155. https://doi.org/10.1111/bjet.12747

Pekrun, R. (2020). Self-report is indispensable to assess students' learning. *Frontline Learning Research*, *8*(3), 185–193. https://doi.org/10.14786/flr.v8i3.637

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist, 37*(2), 91–105. https://doi.org/10.1207/S15326985EP3702_4

Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual Review of Psychology, 54*(1), 547–577. https://doi.org/10.1146/annurev.psych.54.101601.145041

Pike, G. R. (1996). Limitations of using students' self-reports of academic development as proxies for traditional achievement measures. *Research in Higher Education, 37*, 89–114. https://doi.org/10.1007/BF01680043

Prichard, J. S., Bizo, L. A., & Stratford, R. J. (2006). The educational impact of team-skills training: Preparing students to work in groups. *British Journal of Educational Psychology, 76*(1), 119–140. https://doi.org/10.1348/000709904X24564

Rack, O., Zahn, C., spsampsps Mateescu, M. (2018). Coding and counting: frequency analysis for group interaction research. In E. Brauner, M. Boos, spsampsps M. Kolbe (Eds.), *The Cambridge Handbook of Group Interaction Analysis* (pp. 277–294). Cambridge University Press. https://doi.org/10.1017/9781316286302.015

R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/

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

Rosseel, Y., Oberski, D., Byrnes, J., Vanbrabant, L., Savalei, V., Merkle, E., Hallquist, M., Rhemtulla, M., Katsikatsou, M., Barendse, M., & others. (2017). Package 'lavaan'. *Retrieved June*, *17*(1), 2017.

Satorra, A., & Bentler, P. M. (1994). *Corrections to test statistics and standard errors in covariance structure analysis (UCLA Statistics Series 2). Los Angeles: University of California at Los Angeles: Department of Psychology.*

Schaufeli, W. B., Bakker, A. B., & Salanova, M. (2006). The measurement of work engagement with a short questionnaire: A cross-national study. *Educational and Psychological Measurement, 66*(4), 701–716. https://doi.org/10.1177/0013164405282471

Schnaubert, L., & Bodemer, D. (2022). Group awareness and regulation in computer-supported collaborative learning. *International Journal of Computer-Supported Collaborative Learning, 17*(1), 11–38. https://doi.org/10.1007/s11412-022-09361-1

Schoor, C., & Bannert, M. (2011). Motivation in a computer-supported collaborative learning scenario and its impact on learning activities and knowledge acquisition. *Learning and Instruction, 21*(4), 560–573. https://doi.org/10.1016/j.learninstruc.2010.11.002

Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational Psychology Review, 7*(4), 351–371. https://doi.org/10.1007/BF02212307

Shimizu, I., Kikukawa, M., Tada, T., Kimura, T., Duvivier, R., & Van Der Vleuten, C. (2020). Measuring social interdependence in collaborative learning: Instrument development and validation. *BMC Medical Education, 20*(1), 177. https://doi.org/10.1186/s12909-020-02088-3

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

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

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education, 2*, 53.

Tentama, F., & Anindita, W. D. (2020). Employability scale: Construct validity and reliability. *International Journal of Scientific & Technology Research, 9*(4), 3166–3170.

Tumpa, R. J., Skaik, S., Ham, M., & Chaudhry, G. (2022). A holistic overview of studies to improve group-based assessments in higher education: A systematic literature review. *Sustainability, 14*(15), 9638. https://doi.org/10.3390/su14159638

Umemoto, T., Tanaka, K., & Yada, N. (2018). Development of a motivational regulation strategies scale for cooperative learning. *The Japanese Journal of Psychology, 89*(3), 292–301. https://doi.org/10.4992/jjpsy.89.17217

Vriesema, C. C., & McCaslin, M. (2020). Experience and meaning in small-group contexts: fusing observational and self-report data to capture self and other dynamics. *Frontline Learning Research*, *8*(3), 126–139. https://doi.org/10.14786/flr.v8i3.493

West, M. A. (2012). *Effective teamwork: Practical lessons from organizational research*. John Wiley & Sons.

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

Xiong, Y., So, H.-J., & Toh, Y. (2015). Assessing learners' perceived readiness for computer-supported collaborative learning (CSCL): A study on initial development and validation. *Journal of Computing in Higher Education, 27*(3), 215–239. https://doi.org/10.1007/s12528-015-9102-9

Xu, B., Stephens, J. M., & Lee, K. (2024). Assessing student engagement in collaborative learning: Development and validation of new measure in China. *The Asia-Pacific Education Researcher, 33*(2), 395–405. https://doi.org/10.1007/s40299-023-00737-x

Xu, F., & Correia, A.-P. (2024). Measuring mutual engagement in the context of middle-school pair programming: Development and validation of a self-reported questionnaire. *Computers in Human Behavior Reports, 14*, 100415. https://doi.org/10.1016/j.chbr.2024.100415

Zahn, C., Ruf, A., spsampsps Goldman, R. (2021). Video data collection and video analyses in CSCL research. In U. Cress, C. Rosé, A. Wise, spsampsps J. Oshima (Eds.), *International Handbook of Computer-Supported Collaborative Learning*. Springer.

Zheng, L., Long, M., Niu, J., & Zhong, L. (2023a). An automated group learning engagement analysis and feedback approach to promoting collaborative knowledge building, group performance, and socially shared regulation in CSCL. *International Journal of Computer-Supported Collaborative Learning, 18*(1), 101–133. https://doi.org/10.1007/s11412-023-09386-0

Zheng, X.-L., Gu, X.-Y., Lai, W.-H., Tu, Y.-F., Hwang, G.-J., & Wang, F. (2023b). Development of the social metacognition inventory for online collaborative argumentation: Construct validity and reliability. *Educational Technology Research and Development, 71*(3), 949–971. https://doi.org/10.1007/s11423-023-10220-5

Authors and Affiliations

Loris T. Jeitziner[1,3] · Lisa Paneth[1,3] · Oliver Rack[1] · Susanne Bleisch[2] · Carmen Zahn[1]

✉ Loris T. Jeitziner
loris.jeitziner@fhnw.ch

Lisa Paneth
lisa.paneth@fhnw.ch

Oliver Rack
oliver.rack@fhnw.ch

Susanne Bleisch
susanne.bleisch@fhnw.ch

Carmen Zahn
carmen.zahn@fhnw.ch

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

[2] School of Architecture, Civil Engineering, and Geomatics, University of Applied Sciences and Arts Northwestern Switzerland, Muttenz, Switzerland

[3] Faculty of Psychology, University of Basel, Basel, Switzerland