

Article

Self-Assessment of Teamwork Skills Among Adolescents: Psychometric Properties of the Collaborative Skills Scale

Anita Pásztor-Kovács ^{1,*} , Attila Pásztor ^{1,2} , Yong Liu ³  and Gyöngyvér Molnár ^{1,2} 

¹ Institute of Education, University of Szeged, H-6723 Szeged, Hungary; attila.pasztor@edu.u-szeged.hu (A.P.); gymolnar@edpsy.u-szeged.hu (G.M.)

² MTA-SZTE Digital Learning Technologies Research Group, Institute of Education, University of Szeged, H-6723 Szeged, Hungary

³ School of Foreign Languages, Chongqing Technology and Business University, Chongqing 400067, China; yong.liu@ctbu.edu.cn

* Correspondence: pasztor-kovacs@edpsy.u-szeged.hu

Abstract

Monitoring the development of increasingly essential collaborative skills at the individual level within a classroom context requires effective, easy-to-use, and quick measurement tools. These tools should provide global feedback on the skillset rather than reflecting performance in a single group task. A self-rated questionnaire is a beneficial option for this purpose. The aim of our research is to develop a self-rated scale for adolescents, the Collaborative Skills Scale (CoSS), which provides a global assessment of students' teamwork competence. Additionally, using our assessment instrument, we seek to explore what patterns adolescents' self-ratings show to be connected to their collaborative skills. A total of 2128 Grade 8 students participated in our online data collection. The Collaborative Skills Scale was developed based on the collaborative problem-solving model of the ATC21S project. Confirmatory factor analyses yielded a reliable and structurally valid 18-item scale (Cronbach's $\alpha = 0.90$; $\chi^2 = 1802.83$, $df = 132$, $p < 0.01$; CFI = 0.944; TLI = 0.935; RMSEA = 0.077; SRMR = 0.031), which can provide educational practitioners with an effective formative assessment tool for monitoring and supporting the development of teamwork skills. Ideally, it should be utilized in combination with other instruments, such as peer- or teacher-rated scales, to ensure a comprehensive understanding of students' collaborative skills. In line with previous findings, students tended to rate their teamwork skills above average. The implications of this potentially biased self-evaluation among adolescents in terms of collaborative competence are discussed.

Keywords: collaborative skills; teamwork skills; self-rated questionnaire; adolescents; online assessment



Academic Editors: Jeong Hoon Choi and Mian Wang

Received: 29 September 2025

Revised: 14 January 2026

Accepted: 19 January 2026

Published: 21 January 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

The continuous, dynamic development of information and communication technology, along with significant social and economic changes, has led to a radical transformation of labour market needs around the turn of the millennium (Kozma, 2009). The literature often refers to the skills that employees should possess to meet the changing workplace demands as 21st century skills (Binkley et al., 2012; Chai et al., 2024). Over the past two decades, numerous research initiatives have sought to identify and conceptualize these skills, and their findings have informed a wide range of educational frameworks and policy documents (e.g., Commission for the European Communities, 2008; Griffin et al., 2012;

Hao et al., 2017; OECD, 2010, 2017; Voogt & Roblin, 2012; Thornhill-Miller et al., 2023). The ability to collaborate—group work activity that takes place to achieve common goals (Hesse et al., 2015)—plays a central role in all of these documents and frameworks. Recent research examining employer selection criteria has confirmed the increasing importance of teamwork skills (Ferreira et al., 2023; Rios et al., 2020; Tushar & Sooraksa, 2023).

Changing workplace expectations have created new demands in the field of formal education as well. During their school years, students need to acquire skills that were previously less essential in the labour market. Teaching students to work in teams and solve problems collaboratively has become a priority in primary, secondary, and higher education (Chu et al., 2021; Griffin & Care, 2015; Karaca-Atik et al., 2023; Rojas et al., 2021; Roshid & Haider, 2024; Stehle & Peters-Burton, 2019; Tushar & Sooraksa, 2023).

Students with social, emotional, or behavioural difficulties represent a subgroup for whom the development of collaborative skills is particularly important, as they may face disadvantages when entering the job market compared to their typically developing peers (Carter & Lunsford, 2005; Hakkarainen et al., 2016). Therefore, educational practices addressing the needs of these students should place greater emphasis on fostering collaborative skills, contributing to more equitable employment opportunities in adulthood.

Whether we choose to foster teamwork competence indirectly through collaborative learning methods or implement a training program that directly develops collaborative skills, it is essential to continuously gather information throughout the process. This ongoing monitoring makes it possible to identify areas for improvement for both students and educational practitioners, such as teachers, special education teachers, and school psychologists (Csapó et al., 2012; Sortwell et al., 2024). To support this process, efficient formative assessment tools are required.

1.1. The Assessment of Teamwork Skills in a Pedagogical Context

In the last decade, many innovative research approaches have emerged to measure Collaborative Problem Solving (ColPS) skills (Von Davier et al., 2017). For example, these approaches use virtual agents, eye-tracking tools, or analyse conversations by tracking response times and actions to identify behavioral patterns (e.g., Han et al., 2023; Olsen et al., 2017; Stoeffler et al., 2020). However, due to their technological requirements and the expert knowledge needed to evaluate the resulting data, these approaches are not the most feasible options for classroom assessments, where educational practitioners require quick and easy-to-administer tools.

To fulfil the pedagogical goal of monitoring the development of students' collaborative skills, potential instruments need to assess individuals and provide feedback on the overall skillset, rather than reflecting the outcome of a single group task. While teamwork assessment has a strong empirical history (Salas et al., 2017), most measurement tools evaluate collaboration success at the group level, typically by observing the completion of a specific task carried out by the group using various evaluation criteria (Greiff, 2012; Lewis, 2003; Smith-Jentsch et al., 2008). Given the aim of tracking students' collaborative skills at the individual level and providing a comprehensive overview of them in a classroom context, the most beneficial option appears to be the questionnaire method.

For evaluating teamwork skills in an educational environment, usually self-, peer- and teacher-rated questionnaires are utilized (e.g., Hastie et al., 2014; Loughry et al., 2007; Taggar & Brown, 2001; Zhuang et al., 2008). Several pros and cons can be mentioned connected to the three different assessment types. The obvious advantage of peer and teacher evaluation is that the results are less biased by the effect of social desirability. On the other hand, the objectivity of these solutions can be influenced by the personal relationship of the raters and the rated ones (Loughry et al., 2007). Another disadvantage of them is

that schoolmates and teachers face students' teamwork skills exclusively in educational context; they cannot observe their collaborative behaviour in any other social interaction (Zhuang et al., 2008). These latter arguments suggest that self-assessment might be more informative than the other two options. Naturally, in terms of validity, a lot depends on how much self-reflectivity students show, how impartially they can evaluate themselves. The question arises: at what age can their self-judgments be considered acceptably objective?

It seems quite challenging to form a clear perspective on this issue. Although evidence from the Children's Social Desirability Scale shows that social desirability decreases with age (Ng et al., 2025), it is not possible to identify a specific point in adolescence at which this bias becomes negligible and self-perceptions can be considered fully accurate. Rather, the degree of objectivity in adolescents' self-evaluations appears to be highly dependent on the specific construct being assessed.

The literature suggests that social desirability bias tends to be more pronounced in self-report measures when items concern socially sensitive issues or socially normative attributes (Krumpal, 2013). As discussed above, being perceived as a competent team player has become an increasingly valued social characteristic. Ahonen and Kinnunen (2015), in a study of 718 students aged 11 to 15 who evaluated the importance of various 21st-century skills, found that collaboration was ranked as the most important among the 14 skills assessed. These findings suggest that adolescents are well aware of the importance of teamwork skills. However, we currently lack clear evidence regarding the extent to which this norm is internalized, specifically whether not being a capable team player is experienced as socially undesirable, something to be ashamed of, or something to conceal in self-reports.

Consequently, to obtain a clearer understanding of adolescents' self-perceptions related to collaborative competence, further empirical research is needed. To date, most self-report instruments assessing collaborative skills have been developed for later developmental stages, typically for use in young adulthood within higher education contexts or in adulthood within organizational and human resources settings (e.g., Brock et al., 2017; Britton et al., 2017). We identified only two studies that reported on adolescents' self-assessment of teamwork skills (Strom et al., 2024; Zhuang et al., 2008). In the study by Strom et al. (2024), high school students evaluated both their own and their peers' teamwork behaviours across 25 skill statements following a four-week cooperative learning project. The results indicated that students frequently endorsed positive teamwork behaviours for both self- and peer ratings. On average, positive endorsements were observed for 76.4% of the items in self-ratings and 69.3% in peer ratings. Similar tendencies can be observed in the study by Zhuang et al. (2008), which assessed adolescents' teamwork and collaboration skills using a 30-item self-report instrument with a six-point Likert response format. The results indicated that item-level means consistently exceeded the theoretical midpoint of the scale (i.e., 3). When aggregated across items, students' self-ratings were, on average, approximately 1.12 standard deviations above the scale midpoint. These results suggest that adolescents tend to perceive their own teamwork abilities as relatively high.

An additional and equally important question concerns the accuracy of these self-evaluations. Evidence from Strom et al. (2024) suggests a relatively high degree of convergence between self- and peer ratings, with agreement rates ranging from 87% to 99%, indicating that students tended to evaluate both themselves and their peers in a similar manner. However, this convergence may partly reflect the structured context of the assessment, as students received guidance and engaged in explicit discussions about teamwork skills during the cooperative learning project (Strom et al., 2024). Findings reported by Zhuang et al. (2008) provide a more nuanced picture regarding the validity of adolescents' self-assessments. In their study, associations between self-reported teamwork subscales

and a situational judgment test were moderate ($r = 0.47\text{--}0.60$), while correlations with teacher ratings of students' collaborative skills were weaker ($r = 0.14\text{--}0.19$). Taken together, the available empirical evidence is insufficient to draw firm conclusions about the accuracy of adolescents' self-judgments, as the findings present a mixed picture with respect to convergent validity.

To sum up, very few self-report instruments are currently available for assessing adolescents' collaborative skills. Of the two self-report questionnaires cited here, one does not fully meet established psychometric standards with regard to construct validity (Zhuang et al., 2008), while the other is tied to a specific four-week group project (Strom et al., 2024), which limits its suitability for providing global feedback on students' collaborative skills independent of a fixed group composition.

1.2. Theoretical Models for Describing Collaborative Problem Solving—The ATS21C Model of ColPS

The existing literature reveals a clear gap in the availability of psychometrically sound self-rated instruments designed to assess adolescents' collaborative competence. To address this gap, it is essential to develop a new self-rated scale tailored to this age group. The development of such an instrument must be grounded in a well-established theoretical framework.

The OECD's Programme for International Student Assessment (PISA) in 2015, which investigated Collaborative Problem Solving as the fourth area alongside Mathematics, Science, and Reading (OECD, 2017), was followed by many initiatives to measure this construct and the creation of several new theoretical models to describe it. Chai et al. (2024) identified nine models of its structure (e.g., Camara et al., 2015; Liu et al., 2016).

Prior to the PISA 2015 survey, the literature used the terms groupwork, teamwork, collaboration, cooperation, and collaborative problem solving synonymously, referring to a set of social skills (O'Neil et al., 2003). However, the creators of the PISA framework took the interpretation of ColPS in a new direction by building upon their existing problem-solving ability model from the 2012 dynamic problem-solving assessment, extending it with a collaborative dimension (OECD, 2017). This new conceptualization of ColPS as a two-dimensional construct had a significant impact: of the nine theoretical models identified by Chai et al. (2024), eight include cognitive aspects in addition to social ones.

The overall nine models diverse in its content as well as in its elaborateness. In our view, the two-dimensional hierarchical model of the ATC21S project (Hesse et al., 2015, Figure 1) is the most detailed—therefore the most suitable one for building a questionnaire on it. It is beyond our scope to provide a complete, in-depth analysis of Hesse and his colleagues' work on the ColPS construct; instead, we will briefly introduce their key ideas and definitions related to the social component, which is relevant to our research.

ColPS is described as a set of nine social and nine cognitive subskills. The nine social subskills are organized under three core skills: Participation, Perspective taking, and Social regulation. The first core skill, Participation, includes Action, indicated by "activity within the environment"; Interaction, which refers to "interacting with, prompting, and responding to the contributions of others"; and Task completion, with the indicator of "undertaking and completing a task or part of a task individually." The second core skill, Perspective Taking, contains Adaptive responsiveness, expressed in the behaviour of "ignoring, accepting, or adapting contributions of others"; and Audience awareness, which means "the awareness of how to adapt behaviour to increase suitability for others". The third core skill, Social regulation, consists of Negotiation, with the indicator of "achieving a resolution or reaching compromise"; Self-evaluation, which represents "recognizing own strengths and weaknesses"; Transactive memory, which refers to "recognizing strengths and

weaknesses of others”; and Responsibility initiative, which means “assuming responsibility for ensuring parts of task are completed by the group (Table 1)” (Hesse et al., 2015, p. 43).

Rate the sentence below on a scale from 1 to 7 according to how much you agree with it! Answer by clicking!

Number 1 indicates that the statement does not describe you at all, while number 7 indicates that it completely describes you.

When we work in groups, mostly ...

I tell my peers if I hold the opposite opinion.

• 1
• 2
• 3
• 4
• 5
• 6
• 7

1 = Does not describe me at all.
7 = Completely describes me.

Back
Next

Figure 1. Screenshot of the CoSS running in the eDia system (the item examines the Negotiation subskill under the Social regulation subscale).

Table 1. The collaborative problem solving model of Hesse et al. (2015, the table is based on the detailed description of the model provided on pp. 41–52).

Collaborative problem solving	
Social skills	Cognitive skills
Participation	Task regulation
Action	Organization
Interaction	Setting goals
Task completion	Resource management
Perspective taking	Flexibility and ambiguity
Adaptive responsiveness	Collect elements of information
Audience awareness	Systematicity
Social regulation	Learning and knowledge building
Negotiation	Relationships
Self-evaluation	Rules: “If...then”
Transactive memory	Hypothesis: “what if...”
Responsibility initiative	

2. Aims and Research Questions

The increasing relevance of collaborative skills in the labour market underscores the importance of formal education in fostering their systematic development, particularly among students with social, emotional, and behavioural difficulties, to mitigate the risk of future disadvantage in employment. To monitor the development of these skills at the individual level within the classroom context, effective, easy-to-use, and quick-to-

administer measurement tools are necessary. The self-rated questionnaire method appears to be a suitable approach for this purpose. While many self-assessed scales are available to measure collaborative skills in higher education, there is a lack of such instruments for adolescent students. As a result, there is very little data available to describe this age group's self-perception of their teamwork skills. Therefore, the aim of our research is to develop a self-rated questionnaire for adolescents—the Collaborative Skills Scale (CoSS)—which is designed to provide global feedback on students' teamwork competence. To this end, we analyse the psychometric properties of the CoSS and explore how adolescents perceive themselves as collaborators through this instrument. The following research questions are to be investigated:

RQ1: What is the reliability of the CoSS among adolescents?

RQ2: Can the CoSS be considered structurally valid?

RQ3: How do the items and the overall test function across the latent collaborative ability continuum according to item response theory?

RQ4: What patterns do students' self-ratings reveal in relation to their collaborative skills?

3. Method

3.1. Sample and Procedure

A total of 2193 Grade 8 students ($M_{age} = 14.75$ years; $SD = 0.47$) from 135 classes across 84 primary schools participated in the study. Of these, 65 students did not complete the scale and were excluded due to missing data, resulting in a final analytic sample of 2128 students. The final sample included 1117 girls (52.7%), 1003 boys, and eight students who did not report their gender.

The participating schools were drawn from the Hungarian Educational Longitudinal Program (HELP), a nationwide longitudinal research program that follows students' learning and development from Grades 1 to 8 across multiple domains (Csapó, 2014). The sampling design of the program aims to approximate representativeness at the regional, county, and settlement-type levels, and the sample covers approximately 5% of the national student cohort at the target grade level. The school served as the primary sampling unit, and all participating institutions were mainstream primary schools. Although the overall distribution of the sample closely reflects the national population with respect to geographic characteristics, formal goodness-of-fit testing indicated statistically significant differences between the sample and the national distribution ($\chi^2 = 201.29$, $df = 6$, $p < 0.01$). For example, schools located in Budapest, the capital city, were underrepresented. Participation in the program was voluntary: invited schools could decide whether to involve one or multiple eighth-grade classes. For the present data collection, 146 schools were invited, representing approximately 4870 students. Of these, 2128 students completed the survey, resulting in an overall response rate of approximately 44%.

Taken together, the sample provides nationwide coverage and includes students from diverse regional and social backgrounds. While it cannot be considered strictly representative in a statistical sense, it is sufficiently heterogeneous to provide a solid basis for analyses of scale functioning and response patterns in students' self-reports of collaborative skills.

Data collection was conducted in the ICT laboratories of the participating schools using the eDia online assessment system, a nationwide platform supporting students' development across multiple domains through diagnostic assessments and online developmental programs (Csapó & Molnár, 2019; Molnár & Csapó, 2019). The system employs individual measurement identifiers and ensures student anonymity; no personally identifiable information was collected during the assessment.

The study received approval from the university's ethics committee, and all procedures adhered to established ethical guidelines for educational research, including those related to informed consent, privacy, confidentiality, and data management. Schools received detailed information about the study and the data collection procedures, along with written informed consent materials, which were distributed to parents or legal guardians by the participating teachers. Students were informed by their teachers that participation was voluntary and that they could withdraw at any time without any negative consequences.

3.2. Instrument

The CoSS was based on the theoretical model of Hesse et al. (2015). For each of the nine social subskills, four items were developed, with one or two being reverse-scored (Table 2). The scale contained a total of 36 items, 15 of which were reverse-scored (Appendix A). The Perspective taking skill was defined by two subskills, making this subscale the shortest of the three, with four items for the two subskills, totaling 8 statements under Adaptive responsiveness and Audience awareness. In contrast, Social regulation was described by four subskills in the model, so we assigned the largest number of items to this subscale—16 in total. The third subscale, Participation, was measured using 12 items.

Table 2. Sample items from the CoSS in relation to the subscales and their corresponding subskills.

Subscale	Subskill	Number of Items	Sample Item When We Work in Groups, Mostly...
Participation	Action	4	I retreat into the background. (R)
	Interaction	4	I add opinions and comments to the others' suggestions.
	Task completion	4	I don't get to the end of the task assigned to me. (R)
Perspective taking	Adaptive responsiveness	4	I find it difficult to use the suggestions of the others. (R)
	Audience awareness	4	I find the common ground with everybody.
Social regulation	Negotiation	4	I try to reconcile the parties if there is a disagreement.
	Self-evaluation	4	I tell the others if I feel I could do some other part of the work more efficiently.
	Transactive memory	4	it doesn't matter to me who takes which task. (R)
	Responsibility initiative	4	I don't feel disappointed if our group is not successful. (R)

Note: R = Reverse-scored item.

The scale was first piloted in a small-scale study involving 96 Grade 8 students using a paper-based version of the test. Although the reliability indices were promisingly high, the sample size was insufficient to conduct confirmatory factor analysis. Therefore, this initial trial was followed by a second pilot study with a larger sample of 871 Grade 8 students. The results of this second study led to several subsequent revisions of the scale.

The confirmatory factor analysis suggested that more than half of the items, including all reverse-coded items, would need to be removed to obtain acceptable model fit. However, the reliability of the total scale was high, and the reliabilities of the subscales were also satisfactory. For this reason, we retained the full item set for the purposes of the present study. At the same time, we undertook a systematic content review to evaluate whether the items adequately captured the relevant components of the theoretical model, and to identify items with negative or very low correlations with the total scale. This process led to substantial revisions in item wording and formulation, and consequently only eight items remained unchanged in the revised version of the scale.

In the initial version of the scale, students were asked to evaluate the statements on a five-point Likert-type scale indicating the extent to which each statement described them. The majority of respondents rated their collaborative skills as relatively highly developed, resulting in a right-skewed response distribution. In addition, the variance of the composite scores was low relative to their mean. To address these issues, we replaced the five-point

scale with a seven-point scale (1—Does not describe me at all, to 7—Completely describes me), with the expectation that this would enable respondents to provide more differentiated self-assessments and thereby improve the reliability of the instrument.

We also modified the presentation of the items. In the first data collection, 5–6 statements were displayed on a single page; in the present version, only one item was presented at a time, embedded in a template that included the instruction and the seven-point response scale (Figure 1). We judged that this format makes the statements easier to process, more readable due to the larger font size, and that the placement of the radio buttons is more appropriate within the less crowded layout.

3.3. Data Analysis

The statistical analysis of descriptive statistics, reliability indices, item-total correlations, and Pearson's correlation coefficients for the subscales was conducted. To examine structural validity, confirmatory factor analysis (CFA) was performed. Since the observed variables were considered ordinal, the WLSMV (Diagonally Weighted Least Squares) procedure was used. To examine test- and item-level difficulty, item response theory (IRT) was employed, as it provides information on the measurement properties of the scale as well as the fit between the items and students' underlying competencies (Samejima, 1969). With the choice of the parameterization of IRT, partial credit model (PCM) was selected as the analysis approach because of the imbalance of the scale in each item and the indication of some knowledge of the seemingly incorrect responses (Masters, 1982). The PCM was chosen over alternatives such as the Graded Response Model (GRM) also due to its flexibility in accommodating polytomous responses, including our use of reverse-scored items and varying response categories across items. While GRM is also a viable option for Likert scales, PCM provided a more tailored framework for our specific item design.

4. Results

The reliability of the total scale (Cronbach's $\alpha = 0.88$, McDonald's $\omega = 0.86$) and the Participation subscale (Cronbach's $\alpha = 0.82$, McDonald's $\omega = 0.81$) was found to be adequate. The Perspective taking (Cronbach's $\alpha = 0.69$, McDonald's $\omega = 0.67$) and Social regulation (Cronbach's $\alpha = 0.63$, McDonald's $\omega = 0.58$) subscales could also be considered acceptable; however, four items for each of the nine subskills showed low reliability in most cases (Table 3). This indicated that some statements would need to be excluded. Two items were negatively correlated with the total scale. Both of these were reverse-coded items. In the case of the item related to the Responsibility initiative subskill (I don't feel disappointed even if our group does not succeed), the reason for its maladaptive functioning may be the presence of double negatives within the sentence, which could have led to misinterpretation. The other negatively correlating item was associated with the Negotiation subskill (I feel bad when my peers criticize my work). In this case, the negative correlation is more difficult to explain. As it seems unlikely that adolescents particularly welcome criticism, it is more plausible that this item required such a high level of honest self-reflection that students in this age group may not yet possess. The Negotiation subskill also contained another item with a rather low item–total correlation (0.08; I can easily admit when I am wrong). In this case, the problem may again have been the negative phrasing, which could have caused confusion. Together, these two items resulted in the lowest reliability within the Negotiation subscale.

The two negatively correlated items were removed from the scale. Confirmatory factor analysis was conducted with the remaining 34 items. Based on the theoretical model, testing the nine-, three-, and one-dimensional models would all have been justified. However, the four-item variable sets assessing the nine subskills showed such low reliability that

testing the nine-dimensional model did not seem reasonable. Therefore, the one- and three-dimensional models were tested. Both the one-dimensional ($\chi^2 = 24326.04$; $df = 527$; $p < 0.01$; CFI = 0.570; TLI = 0.542; RMSEA = 0.146) and the three-dimensional model ($\chi^2 = 2427.22$; $df = 524$; $p < 0.01$; CFI = 0.571; TLI = 0.541; RMSEA = 0.146) showed poor fit indices.

Table 3. Means, standard deviations and reliability indices of the 36-item scale and its subscales.

Scale	Nr. of Items	Cronbach- α	McDonald's ω	Mean (SD)	Theoretical Min.–Max.
Total scale	36	0.88	0.86	167.04 (25.58)	36–252
Participation	12	0.82	0.81	58.23 (11.35)	12–84
Action	4	0.74	0.75	19.82 (5.06)	4–28
Interaction	4	0.64	0.66	19.68 (4.43)	
Task completion	4	0.47	0.43	18.73 (4.00)	
Perspective taking	8	0.69	0.67	37.98 (7.13)	8–56
Adaptive responsiveness	4	0.38	0.39	18.41 (3.62)	4–28
Audience awareness	4	0.63	0.65	2.74 (4.55)	
Social regulation	16	0.63	0.58	7.83 (1.23)	16–112
Negotiation	4	0.16	0.34	18.48 (3.62)	4–28
Self-evaluation	4	0.38	0.46	18.25 (3.80)	
Transactive memory	4	0.44	0.45	17.12 (4.05)	
Responsibility initiative	4	0.33	0.37	16.98 (3.83)	

To improve model fit, item reduction was conducted in a stepwise and theory-informed manner. At each step, the item with the lowest standardized factor loading was removed, while simultaneously considering the underlying structure of the three-dimensional model. Items with standardized factor loadings below 0.40 were considered for deletion in line with commonly used psychometric guidelines (e.g., Cheung et al., 2024). One exception was the item *Sr_nego_v6* (factor loading = 0.32), which was retained to preserve the proportional representation of items across the three dimensions, consistent with the structure of the original 36-item instrument. After eliminating 16 statements, we arrived at an 18-item variable set with adequate fit indices for both the one-dimensional and three-dimensional models (one-dimensional: $\chi^2 = 1875.91$, $df = 135$, $p < 0.01$; CFI = 0.942; TLI = 0.934; RMSEA = 0.078; SRMR = 0.031; three-dimensional: $\chi^2 = 1802.83$, $df = 132$, $p < 0.01$; CFI = 0.944; TLI = 0.935; RMSEA = 0.077; SRMR = 0.031). While the fit of the three-dimensional model proved to be significantly better than that of the one-dimensional model ($\chi^2 = 86.638$, $df = 3$, $p < 0.01$), the changes in approximate fit indices were minimal ($\Delta CFI = 0.002$; $\Delta RMSEA = 0.001$). In addition, strong correlations among the three latent factors in the three-dimensional model further indicate that the one-dimensional model is also justified (Figure 2).

In summary, a total of 18 statements were removed, including all 15 reverse-scored items. The shortened 18-item scale includes six items in the Participation subscale, four items in the Perspective taking subscale, and eight items in the Social regulation subscale (Table 4). At least one of the 18 statements is linked to each of the nine subskills. The Negotiation subskill is assessed by three items, while seven other subskills are assessed by two items each. Self-evaluation is the only subskill tied to which only one item remained in the shortened scale. The reliability of the 18-item scale is high (Cronbach's $\alpha = 0.90$, McDonald's $\omega = 0.90$), and the reliability for the three subscales is also satisfactory, considering the number of items (both Cronbach's α and McDonald's $\omega \geq 0.70$) (Table 5).

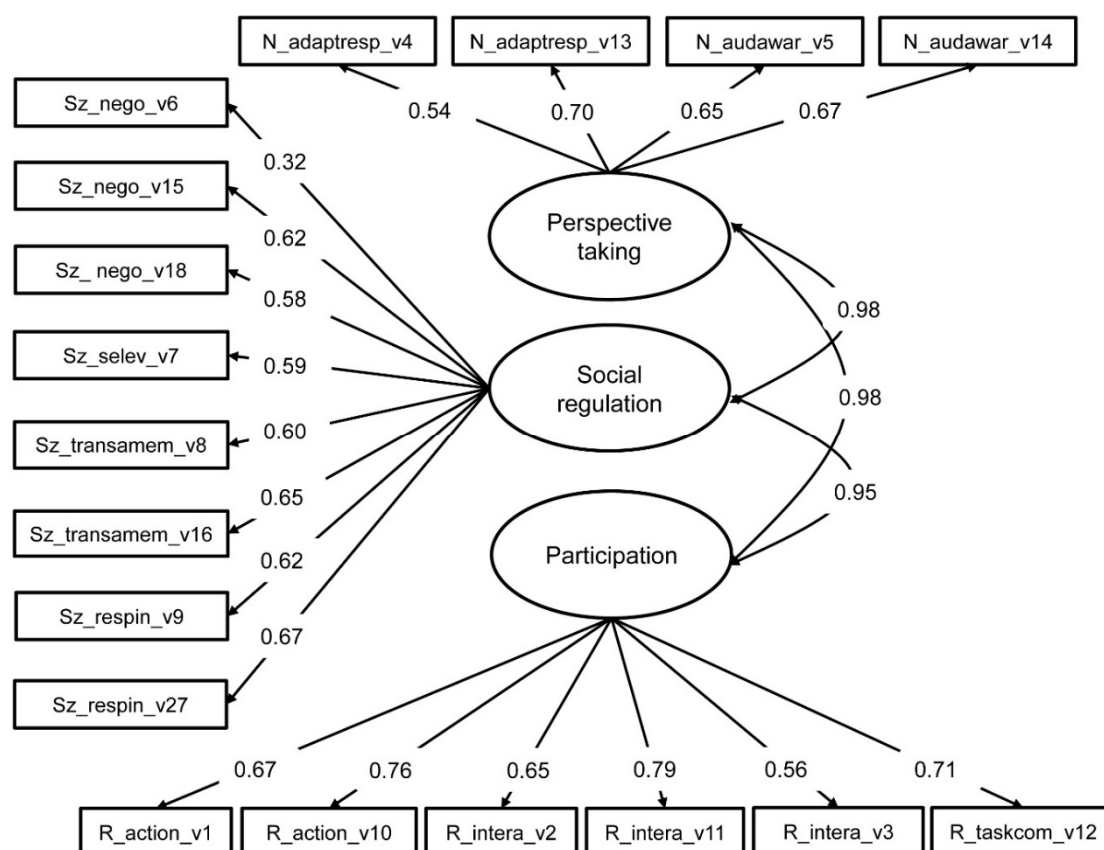


Figure 2. The result of the confirmatory factor analysis on the remaining 18 items of the scale. (The values on the arrows indicate the factor loadings, and the values on the double-ended arrows represent the correlation coefficients between the latent factors).

Table 4. The 18 items of the CoSS scale.

Item	When We Work in Groups, Mostly...
Participation (6 items)	
P_action_v1	I take part actively in the work.
P_action_v10	I am engaged.
P_intera_v2	I react to the others' ideas and suggestions (e.g., with approval or with questions).
P_intera_v11	I share my ideas and thoughts with my peers.
P_taskcom_v3	I keep trying to complete my part of the task until I succeed.
P_taskcom_v12	I try another strategy to solve my task if the previous one did not work.
Perspective taking (4 items)	
Pt_adaptresp_v4	I come up with a good idea while I am listening to my peers' suggestions.
Pt_adaptresp_v13	I develop my peers' ideas further.
Pt_audawar_v5	I find the common ground with everybody.
Pt_audawar_v14	I can explain my ideas so that everybody would understand them.
Social regulation (8 items)	
Sr_nego_v6	I can easily acknowledge if I am wrong.
Sr_nego_v15	I try to reconcile the parties if there is a disagreement.
Sr_nego_v18	I tell my peers if I hold the opposite opinion.
Sr_selfev_v7	I know what job I am the most suitable for.
Sr_transamem_v8	I make suggestions who and what task should do according to what they are skillful in.
Sr_transamem_v16	I indicate if I believe that the division of work should be changed.
Sr_respin_v9	I pay attention to how my groupmates get on with their work.
Sr_respin_v17	I help my peers when I have completed my job.

Table 5. Means, standard deviations and reliability indices of the CoSS and its three subscales.

Scale	Nr. of Items	Cronbach- α	McDonald's ω	Total Score Mean (SD)	Theoretical Min.–Max.
Total scale	18	0.90	0.90	88.46 (16.91)	18–126
Participation	6	0.82	0.82	30.23 (6.55)	6–42
Perspective taking	4	0.70	0.70	19.66 (4.27)	4–28
Social regulation	8	0.76	0.77	38.57 (7.74)	8–56

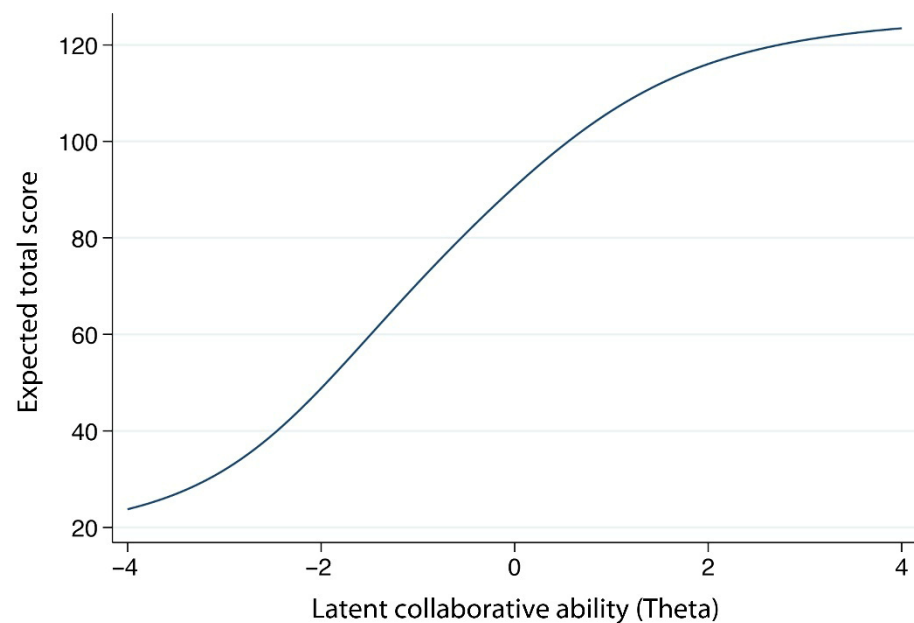
Consistent with the acceptable fit of both the three-dimensional and unidimensional models in the confirmatory factor analysis, the subscales showed moderately to strongly correlated manifest scores, with correlations ranging from $r = 0.72$ to 0.75 , and even higher correlations with the total scale ($r = 0.86$ – 0.93 ; Table 6).

Table 6. Correlation coefficients between the 18-item total scale and its subscales.

Scale	Total Scale	Participation	Perspective Taking
Participation	0.92		
Perspective taking	0.87	0.74	
Social regulation	0.93	0.75	0.72

Note: All correlation coefficients are significant at the $p < 0.01$ level.

The IRT results demonstrated the test and item discrimination parameters of the CoSS. Regarding test item difficulty, based on the test characteristic curve (Figure 3), as students' collaborative ability increases, the expected total score also rises. This indicates that the whole test accords with common sense and effectively measures students' collaborative skills.

**Figure 3.** Test characteristic curve of CoSS.

The test information function represents the summed information provided by all items at a given ability level, with higher values reflecting greater measurement precision (Figure 4). The peak of the test information function curve occurs around $\Theta = -1.5$, suggesting that the scale provides the most information and achieves the highest measurement precision for students with below-average levels of collaborative skills.

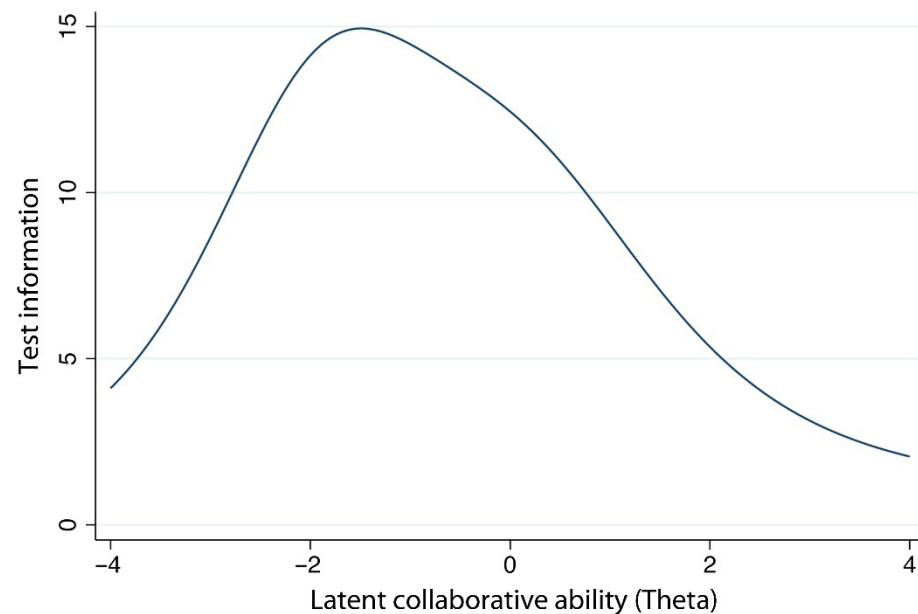


Figure 4. Test information function curve of CoSS. The y -axis represents the total test information, calculated as the sum of item information functions.

Looking more closely at item discrimination parameters in this test, the item information functions for all items are shown in Figure 5. All peaks fall within an information range of 0.72 to 0.95, suggesting a relatively homogeneous contribution of the items to measurement precision. This pattern indicates that no single item disproportionately dominates the information provided by the scale, and that the retained items function with comparable discriminatory power within their most informative ability ranges. Most item peaks between $\text{Theta} = -1.9$ and $\text{Theta} = -0.8$, which aligns with the overall test information trend. In the present sample, approximately 16.4% of students fall within this ability range, indicating that the scale achieves its highest measurement precision for a relatively limited subgroup of students. At the same time, the right-skewed shape of the item information curves suggests that measurement precision does not decline sharply outside the peak region (see Appendix B for the item information functions of the subscales). Rather, the scale continues to provide a meaningful amount of information for students whose ability levels are close to the average, supporting its use beyond the narrow range of maximum information.

Regarding the frequency distribution of responses, students most frequently selected the value of four, followed by notable frequencies for values five, six, and seven (Table 7). Nevertheless, responses were not exclusively concentrated at the upper end of the scale, as lower response values were also chosen by a smaller but meaningful proportion of students. To further contextualize these patterns, we examined the distribution of total scale scores across quartiles. When dividing the total scale into four quartiles, 26.7% of students fell into the top quartile (total scores between 5.5 and 7.0), while 55.5% were classified into the third quartile (total scores between 4.0 and 5.5). A smaller proportion of students scored in the second quartile (16.8%, total scores between 2.5 and 4.0), and only 1.0% were located in the lowest quartile (total scores between 1.0 and 2.5). In addition, when comparing the observed mean score to the theoretical midpoint of the scale (i.e., 4 on the seven-point scale), students' self-ratings were, on average, approximately 0.60 standard deviations above the midpoint, indicating a tendency toward higher-than-average self-evaluations. Taken together, these results indicate that the majority of students perceive their collaborative competence as average or above average.

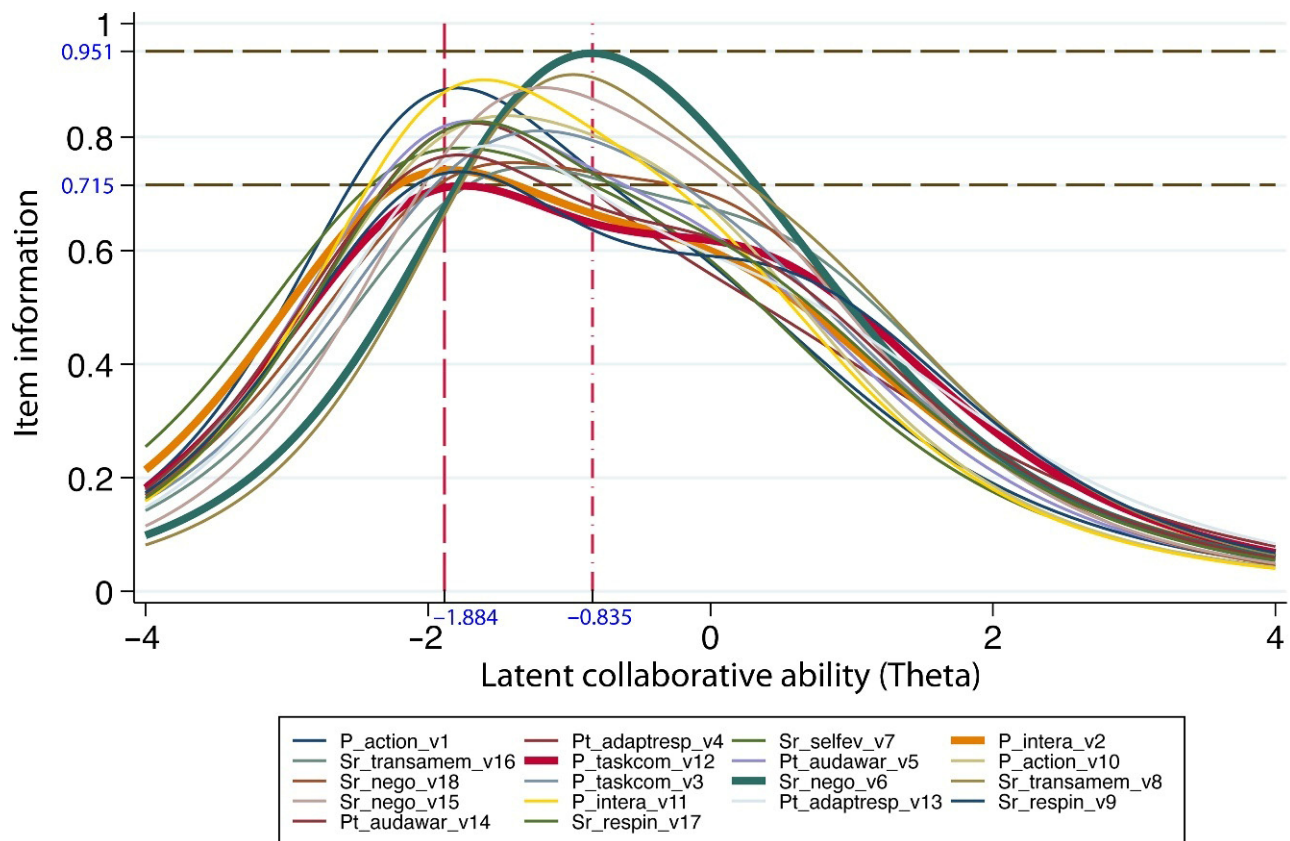


Figure 5. Item information functions of CoSS.

Table 7. Frequency distribution of responses on the 18-item CoSS, along with mean and SD values of the scores associated with each item.

Item	Distribution of the Frequency of Responses (%) 1 = Does Not Describe You at All 7 = Completely Describes You							Mean (Score)	SD (Score)
	1	2	3	4	5	6	7		
P_action_v1	2.2	2.3	6.1	19.5	22.0	25.1	22.7	5.23	1.45
P_action_v10	2.3	3.6	7.2	22.9	17.5	22.1	24.4	5.14	1.54
P_intera_v2	1.9	2.8	8.5	24.7	22.0	21.9	18.2	5.01	1.45
P_intera_v11	2.8	2.5	6.9	21.4	19.0	21.8	25.6	5.19	1.54
P_taskcom_v3	3.1	4.7	9.8	22.3	21.2	21.1	17.8	4.88	1.56
P_taskcom_v12	2.7	3.3	10.7	28.8	20.8	18.8	15.0	4.78	1.48
Pt_adaptresp_v4	2.7	3.2	8.3	22.2	25.7	24.1	13.8	4.93	1.44
Pt_adaptresp_v13	3.4	4.2	9.4	27.3	23.4	20.1	12.4	4.73	1.48
Pt_audawar_v5	2.5	2.8	7.5	22.9	21.1	21.7	21.6	5.09	1.50
Pt_audawar_v14	2.4	3.1	8.3	27.8	20.3	20.5	17.6	4.92	1.48
Sr_nego_v6	6.3	6.2	11.2	21.2	18.6	19.4	17.2	4.66	1.73
Sr_nego_v15	4.9	4.1	9.9	23.7	18.7	18.5	20.2	4.83	1.66
Sr_nego_v18	2.7	4.3	9.6	26.1	19.8	17.0	20.5	4.89	1.56
Sr_selfev_v7	1.3	2.6	6.3	18.8	21.1	24.5	25.5	5.31	1.43
Sr_transamem_v8	7.8	6.1	10.2	25.7	20.2	15.3	14.8	4.50	1.72
Sr_transamem_v16	4.0	5.4	11.7	27.2	21.1	15.8	14.7	4.62	1.57
Sr_respin_v9	2.8	3.4	9.2	29.8	23.3	16.9	14.7	4.77	1.46
Sr_respin_v17	2.7	3.4	7.6	24.2	22.3	20.4	19.4	4.99	1.50

5. Discussion

The aim of this study was to develop an easy-to-use and quick-to-administer self-report scale for assessing collaborative skills that can facilitate the monitoring and thereby the fostering of teamwork competencies, which have growing relevance in contemporary work environments (Ferreira et al., 2023; Rios et al., 2020; Tushar & Sooraksa, 2023). Their development thus represents an important goal of formal education, especially for students with social, emotional, and behavioural difficulties. These students may face disadvantages in functioning as socially competent members of a team, making targeted support particularly important for ensuring more equal opportunities in their future working lives (Carter & Lunsford, 2005; Hakkarainen et al., 2016). Furthermore, we aimed to explore adolescents' self-evaluative judgments of their collaborative skills. To this end, we formulated four research questions.

Research question 1 addressed the reliability of the CoSS among adolescent students. The results of the item-total correlation and confirmatory factor analysis produced an 18-item scale that demonstrates good reliability. Cronbach's α and McDonald's ω values are satisfactory across all subscales, even for the shortest one, Perspective taking, which consists of only four items.

Research question 2 investigated the structural validity of the CoSS. The model fits for both the one- and three-dimensional models support the theory of Hesse et al. (2015), which posits that social skills are organized as a large unit consisting of three main skills. The nine subskills associated with the three main skills are represented in the 18-item scale, with at least one item per subskill. While this limited number of items does not allow for a comprehensive assessment of all nine subskills, the inclusion of at least one item addressing each subskill provides support for the scale's content validity. Between the one- and three-dimensional approaches, the three-dimensional model proved to be more suitable for describing the construct; however, although the difference in fit compared to the one-dimensional model was statistically significant, its magnitude was modest. This aligns with the finding that the three subscales are highly correlated at both the latent and manifest levels, suggesting that the unidimensional approach also has strong relevance. Nevertheless, the manifest-level correlations between the subscales are not so high as to imply they measure exactly the same thing; rather, they appear to provide distinct information. Accordingly, in addition to analysing the total scale, we recommend examining the subscales, as they may contribute to a more differentiated understanding of an adolescent's skill level. To support the interpretation of the CoSS scores and its three subscales, we include a scoring guide in Appendix C following the final 18-item version of the scale.

Referring to research question 3, the IRT analyses provided additional evidence regarding the functioning of the CoSS and the ability range in which the instrument operates most effectively. The test characteristic curve confirmed that higher levels of latent collaborative ability are associated with higher expected test scores, indicating that the scale behaves in a theoretically coherent manner and is sensitive to differences in students' collaborative skills across the ability continuum. The test information function further revealed that the CoSS achieves its highest measurement precision at below-average levels of collaborative ability. This finding suggests that the instrument is particularly well suited for identifying and differentiating among students who may experience greater difficulties in collaborative problem solving, which is especially relevant for diagnostic and formative assessment purposes for students with social, emotional and behavioural difficulties. Importantly, although the peak information is concentrated in a relatively narrow ability range, the right-skewed shape of both the test and item information functions indicates that measurement precision does not decline sharply outside this optimal region. Rather, the CoSS

continues to provide a meaningful amount of information for students with ability levels approaching the average. This pattern suggests that, while the scale is most precise for students with lower levels of collaborative skills, it is also capable of assessing students with average levels of ability with adequate measurement precision, though with reduced accuracy at higher ability levels.

Research question 4 examined the pattern of students' self-ratings related to their collaborative skills. We found a pattern consistent with previous research (Strom et al., 2024; Zhuang et al., 2008), in that students tended to rate their collaborative skills above the average. In the study by Zhuang et al. (2008), students' self-ratings were, on average, approximately 1.12 standard deviations above the theoretical midpoint of the scale, indicating a pronounced tendency toward elevated self-evaluations. In the present study, this deviation was more moderate, with students' self-ratings averaging 0.60 standard deviations above the midpoint. This difference suggests that the tendency toward overestimation of collaborative skills may have been less pronounced in our study. One possible explanation lies in differences in response format: the broader range of the seven-point Likert scale used in the present study may have allowed students to express more nuanced judgments of their collaborative competence. However, this interpretation should be treated with caution, as differences in sample characteristics and study context may also have contributed to the observed pattern. Given the descriptive nature of the available data, firm conclusions regarding the sources of these differences cannot be drawn.

The tendency toward relatively elevated self-ratings of collaborative skills may be interpreted optimistically if viewed as an indication that adolescents recognize the growing importance of collaborative competence and seek to align with this expectation by providing responses they believe are socially valued. An alternative explanation is that a substantial proportion of adolescents may hold overly optimistic or insufficiently differentiated self-perceptions of their collaborative skills. Importantly, these interpretations are not mutually exclusive, and both heightened social desirability and limitations in self-reflective accuracy may jointly contribute to the observed pattern of self-ratings.

Limitations and Future Research

One of the major limitations of the present study, which should be addressed in future research, is the lack of external validity evidence. The current study examined only internal structure evidence as sources of validity, with particular emphasis on reliability and factor-analytic results. However, in the absence of comparisons with related external measures, such as teacher- or peer-reported questionnaires or observational indicators derived from students' performance in group tasks, it remains unclear to what extent students' self-ratings can be considered reliable indicators of their collaborative competence. Moreover, participants' tendency to rate themselves above average on the assessment instrument underscores the need to interpret the findings with caution and critical reflection.

Understanding the tendency for adolescents to provide somewhat biased self-ratings of their collaborative skills is also among the aims of our future research. A key question is whether this bias primarily reflects social desirability or whether adolescents genuinely hold an inflated or distorted self-concept regarding their teamwork abilities. If the latter is the case, further questions arise concerning the underlying mechanisms: why do adolescents perceive themselves as more competent collaborators than external assessments might suggest? For instance, a further aim of future research could be to examine the extent to which adolescents' self-ratings of collaborative skills are shaped by broader cultural and educational contexts. Classroom norms regarding cooperation, competition, and feedback may influence how students evaluate their own collaborative competence. In addition, teachers' instructional practices and training related to collaborative learning and

assessment may shape students' understanding of what constitutes effective collaboration and, consequently, how they position themselves on self-report instruments. Cross-cultural comparisons could further clarify whether the tendency toward relatively elevated self-ratings observed in the present study reflects context-specific norms or more general developmental patterns.

In addition, future studies should examine measurement invariance across relevant subgroups, such as gender or other student characteristics. Establishing invariance would be essential both for strengthening the validity evidence of the scale and for ensuring that comparisons between different groups of students are meaningful and unbiased. Without such analyses, it remains uncertain whether the factor structure and item functioning operate equivalently across subpopulations, which limits the interpretability of subgroup differences in collaborative competence.

It should also be noted that, since only 14–15-year-old students participated in the study, our results cannot be fully generalized to the entire adolescent population. Further research should involve both younger (ages 11–13) and older adolescents (ages 16–18) in order to develop an instrument whose results are more broadly generalizable across the full adolescent age range.

Acquiescence bias represents a particularly serious threat to the validity of self-report instruments among adolescents. Using the Big Five Inventory, a self-report questionnaire comparable to the CoSS, [Soto et al. \(2008\)](#) found that, among individuals aged 10 to 20, the influence of acquiescence bias gradually decreases with age; however, around age 14, it can still exert a substantial effect. To address this problem, we included reverse-coded items in our questionnaire; however, all of these items were eliminated during the confirmatory factor analysis. The necessity of removing reverse-coded items in self-report questionnaires administered to adolescents has also been reported by other authors, including in studies examining collaborative skills ([Józsa & Morgan, 2017](#); [Zhuang et al., 2008](#)). It is likely that, similar to their findings, we also encountered the problems of superficial reading, weak reading comprehension, and the aforementioned acquiescence bias. Accordingly, one limitation of the present instrument is that, due to the absence of reverse-coded items, it does not allow for controlling acquiescence bias. This limitation further reinforces our recommendation to complement self-report data with additional sources, such as teacher or peer ratings and classroom observations, in order to obtain a more differentiated picture of students' collaborative skills.

Lastly, the test characteristic curve and the test information function curve indicated that the CoSS is most effective in assessing students with lower to average levels of collaborative competence. While this pattern supports the instrument's usefulness for diagnostic and formative purposes, it also highlights an important direction for future scale development: the inclusion of items that more effectively capture higher levels of collaborative competence. Such an extension would serve two related purposes. It would improve measurement precision for students with higher levels of collaborative skills and enable the detection of developmental change among students who already score relatively high at earlier measurement points. [Hesse et al. \(2015, p. 43\)](#), in their model, describe behavioural indicators associated with low, medium, and high levels of collaborative skills. When extending the scale with items targeting higher competence levels, it would be appropriate to draw on indicators corresponding to the high-level category. For example, a potential new item related to the Interaction subskill could be: I encourage my groupmates to interact with the rest of the group. Another example, linked to the Negotiation subskill, could be: I achieve the resolution of differences within the group.

6. Conclusions

Our research resulted in a new, reliable self-report questionnaire for adolescents that addresses a highly demanded skill set in the 21st century labour market: teamwork skills. The Collaborative Skills Scale demonstrated solid content validity, as it includes items representing both the three core skills and the nine subskills of Hesse et al.'s (2015) theoretical model. Rather than focusing on performance in a single group task, the CoSS provides global feedback on students' collaborative competence, offering educational practitioners such as teachers, special education teachers or school psychologists a formative assessment tool suitable for monitoring and supporting the development of teamwork skills. The CoSS can be administered quickly and easily in a classroom setting, either in paper-based or online form. However, despite its advantages, users should keep in mind that, as a self-rated scale, the results cannot be regarded as objective. Therefore, educators are encouraged to combine the CoSS with additional methods, such as peer rating scales or direct observations of teamwork behaviour, to obtain a more comprehensive understanding of students' collaborative skills. Based on the IRT analysis, expanding the scale with items that assess higher levels of collaboration would be beneficial. Furthermore, our findings contribute to the understanding of adolescents' self-perception regarding collaborative competence: in line with previous research, students tended to rate their teamwork skills above average.

Author Contributions: Conceptualization, A.P.-K. and A.P.; methodology, A.P.-K., A.P. and Y.L.; software, G.M.; validation, A.P., Y.L. and G.M.; formal analysis, A.P.-K., A.P., Y.L. and G.M.; investigation, A.P.-K., A.P., Y.L. and G.M.; resources, G.M.; data curation, A.P.; writing—original draft preparation, A.P.-K., A.P. and Y.L.; writing—review and editing, A.P.-K., A.P., Y.L. and G.M.; visualization, A.P. and Y.L.; supervision, G.M.; project administration, A.P.-K., A.P. and G.M.; funding acquisition, G.M., A.P. and Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: The research was supported by the Hungarian Academy of Sciences Research Programme for Public Education Development grant (KOZOKT2025-4), the Hungarian National Research, Development and Innovation Fund (OTKA K152413), the Research Start-up Project The Assessment and Measurement of Undergraduates' Critical Thinking Disposition for Introduced High-Level Talent of Chongqing Technology and Business University (Grant Number: 2555003) and the University of Szeged Open Access Fund (Grant ID: 8177). Attila Pásztor was additionally supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences (BO/00093/23/2).

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of the DOCTORAL SCHOOL OF EDUCATION, UNIVERSITY OF SZEGED (14/2019, 15 May 2019).

Informed Consent Statement: Parental consent was obtained for the participation of the students in the study.

Data Availability Statement: The original data presented in the study are openly available in [Zenodo] at [<https://doi.org/10.5281/zenodo.13850614>].

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

List of the 36 Items

The items marked with an * are also included in the final 18-item scale.

The items marked with an (R) are reverse-coded items.

All items begin with "When we work in groups, mostly..."

Participation	
Action	I take part actively in the work. * I leave it to the others. (R) I stay in the background. (R) I am engaged. *
Interaction	I keep my thoughts to myself. (R) I react to the others' ideas and suggestions (e.g., with approval or with questions). * I share my ideas and thoughts with my peers. * I comment on other people's ideas.
Task completion	I keep trying to complete my part of the task until I succeed. * I find it hard to come up with a new strategy for my part if the previous one does not work. (R) I try another strategy to solve my task if the previous one did not work. * I do not get to the end of the task I was given. (R)
Perspective taking	
Adaptive responsiveness	I come up with a good idea while I am listening to my peers' suggestions. * I consider other people's ideas, but I do not add any suggestions. (R) I develop my peers' ideas further. * I find it hard to use other people's suggestions. (R)
Audience awareness	I find it hard to make myself understood by my groupmates. (R) The others do not understand my explanation. I find the common ground with everybody. * I can explain my ideas so that everybody would understand them. *
Social regulation	
Negotiation	I can easily acknowledge if I am wrong. * I try to reconcile the parties if there is a disagreement * I feel bad when my peers criticize my work (R) I tell my peers if I hold the opposite opinion. *
Self-evaluation	I know what job I am the most suitable for. * I cannot really judge on my own how well I am doing. (R) I find it hard to decide what kind of task suits me. (R) I speak up if I feel that I could do another part of the work more effectively.
Transactive memory	I make suggestions who and what task should be done according to what they are skillful in. * I indicate if I believe that the division of work should be changed. * It does not matter to me who takes which role. (R) I find it hard to judge who would be suitable for which task. (R)
Responsibility initiative	I don't feel disappointed even if our group does not succeed, (R) I experience it as my own failure if we do not reach our goal. I pay attention to how my groupmates get on with their work. * I help my peers when I have completed my job. *

Appendix B

Item Information Functions of the CoSS Subscales

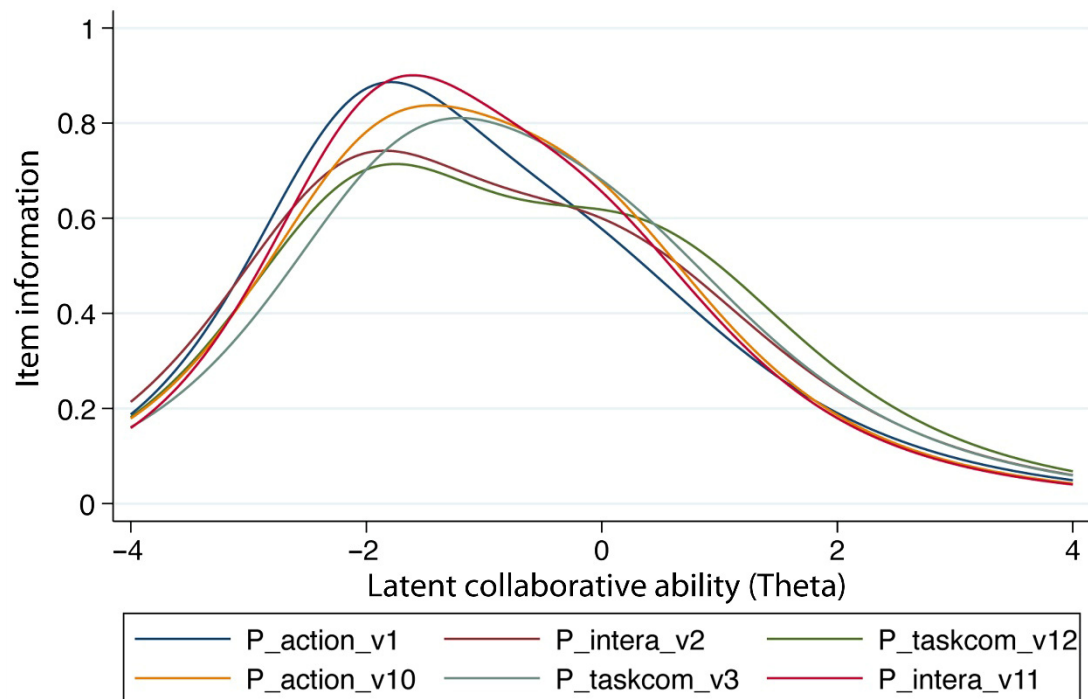


Figure A1. Item information functions of Participation.

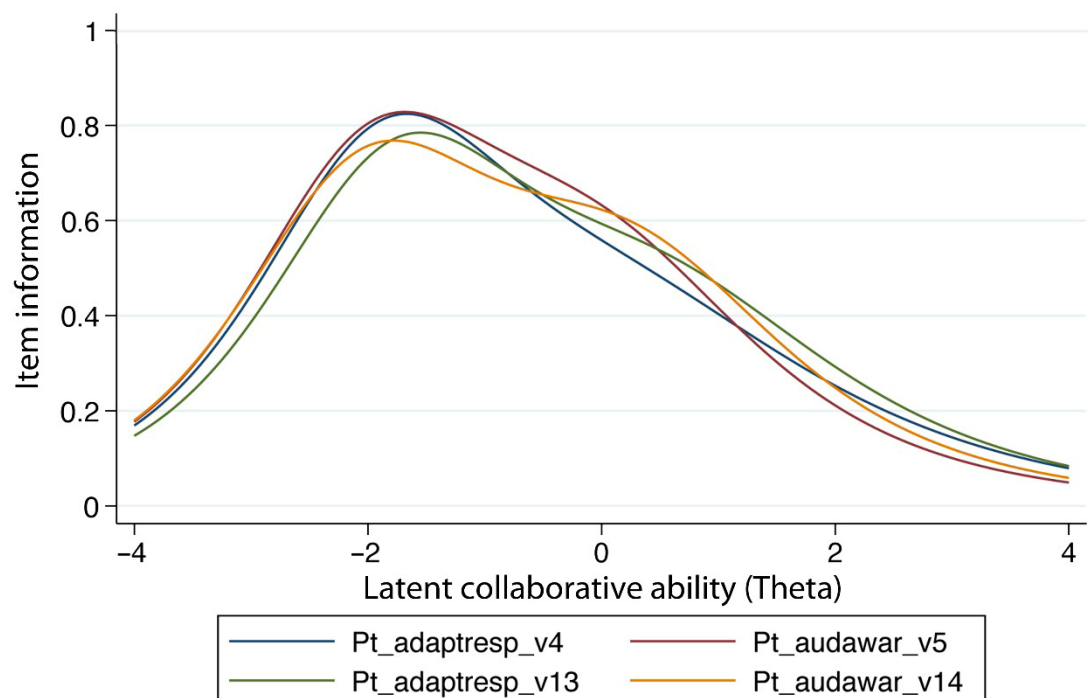


Figure A2. Item information functions of Perspective taking.

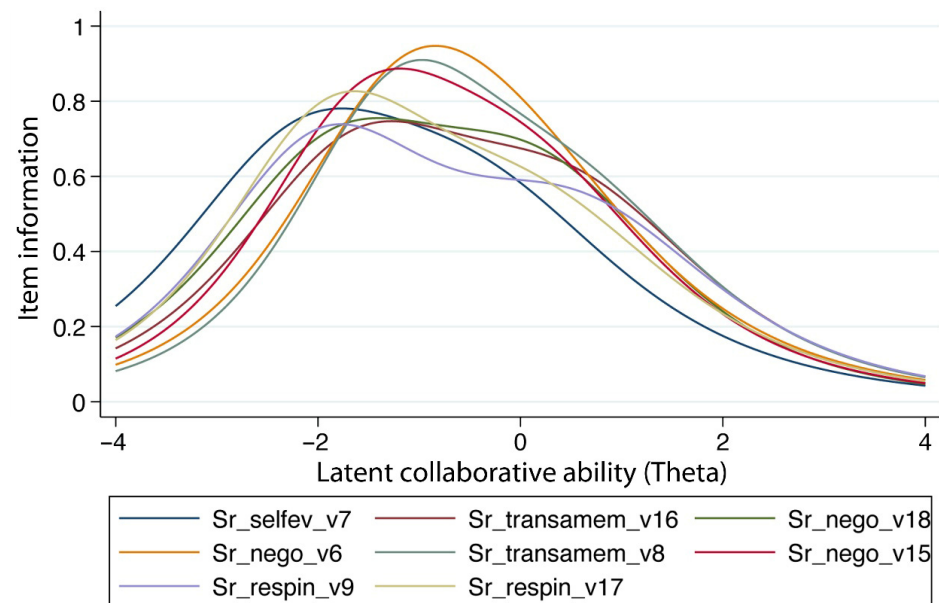


Figure A3. Item information functions of Social regulation.

Appendix C

Appendix C.1. The Paper-Based Version of the Collaborative Skills Scale (CoSS)

The scale below investigates how you usually behave when you do a task in a group. There are no right or wrong answers; you can be honest.

Please recall situations when you had to do an in- or outside-class task or project-work in pairs or in groups. Then, try to decide how much these statements describe you in these situations.

Rate the sentences below on a scale from 1 to 7 according to how much you agree with them and circle the appropriate number. Number 1 indicates that the statement does not describe you at all, while number 7 indicates that it completely describes you. Therefore, the higher the score you mark, the better the statement describes you.

Each sentence starts like this:

When we work in groups, mostly...

1.	I take part actively in the work.	1	2	3	4	5	6	7
2.	I react to the others' ideas and suggestions (e.g., with approval or with questions).	1	2	3	4	5	6	7
3.	I keep trying to complete my part of the task until I succeed.	1	2	3	4	5	6	7
4.	I come up with a good idea while I am listening to my peers' suggestions.	1	2	3	4	5	6	7
5.	I find the common ground with everybody.	1	2	3	4	5	6	7
6.	I can easily acknowledge if I am wrong.	1	2	3	4	5	6	7
7.	I know what job I am the most suitable for.	1	2	3	4	5	6	7
8.	I make suggestions who and what task should be done according to what they are skillful in.	1	2	3	4	5	6	7
9.	I pay attention to how my groupmates get on with their work.	1	2	3	4	5	6	7
10.	I am engaged.	1	2	3	4	5	6	7
11.	I share my ideas and thoughts with my peers.	1	2	3	4	5	6	7
12.	I try another strategy to solve my task if the previous one did not work.	1	2	3	4	5	6	7
13.	I develop my peers' ideas further.	1	2	3	4	5	6	7
14.	I can explain my ideas so that everybody would understand them.	1	2	3	4	5	6	7
15.	I try to reconcile the parties if there is a disagreement.	1	2	3	4	5	6	7
16.	I indicate if I believe that the division of work should be changed.	1	2	3	4	5	6	7
17.	I help my peers when I have completed my job.	1	2	3	4	5	6	7
18.	I tell my peers if I hold the opposite opinion.	1	2	3	4	5	6	7

Note: The original language of the scale is Hungarian. The 18-item final scale was translated from Hungarian to English by two independent native Hungarian-speaking English teachers. The final version published here was prepared by reconciling the two translations, in an effort to reproduce the meaning of the original items as closely as possible.

Appendix C.2. Scoring Guide

The items of the Collaborative Skills Scale provide feedback on how socially competent students are in a group problem-solving situation and how effectively they can collaborate with their peers. Students are asked to rate a series of statements on a seven-point scale according to how much each statement applies to them. The questionnaire consists of three subscales.

The Participation subscale characterizes the student in terms of how much they engage in the task, whether they respond to their peers' initiatives, initiate interactions themselves, and how much effort they invest in completing a given group task.

The Perspective Taking subscale provides feedback on the extent to which the student is open to their peers' contributions, including whether they reject, accept, or build on their suggestions. It also examines whether the student is able to shape their communication so that others can understand it, and whether they are receptive to their groupmates' needs.

The Social Regulation subscale measures the student's ability to make compromises; to recognize their own strengths and weaknesses and take them into account when choosing their role in the group; to understand their peers' abilities and consider them when tasks are distributed; and to feel personal responsibility for ensuring that the group completes its work.

To score the questionnaire, add up the numbers circled by the respondent. The higher the total score, the more developed the student's collaborative skills are. It is also worth considering the scores obtained on each subscale, as they provide feedback on the specific skill areas described above.

The items belonging to each subscale are as follows:

Participation: 1, 2, 3, 10, 11, 12

Perspective Taking: 4, 5, 13, 14

Social Regulation: 6, 7, 8, 9, 15, 16, 17, 18

References

- Ahonen, A. K., & Kinnunen, P. (2015). How do students value the importance of twenty-first century skills? *Scandinavian Journal of Educational Research*, 59(4), 395–412. [\[CrossRef\]](#)
- Binkley, M., Erstad, O., Herman, J., Raizen, S., Martin, R., Miller-Ricci, M., & Rumble, M. (2012). Defining twenty-first century skills. In P. Griffin, B. McGaw, & E. Care (Eds.), *Assessment & teaching of 21st century skills* (pp. 17–66). Springer. [\[CrossRef\]](#)
- Britton, E., Simper, N., Leger, A., & Stephenson, J. (2017). Assessing teamwork in undergraduate education: A measurement tool to evaluate individual teamwork skills. *Assessment & Evaluation in Higher Education*, 42(3), 378–397. [\[CrossRef\]](#)
- Brock, S. E., McAliney, P. J., Ma, C. H., & Sen, A. (2017). Toward more practical measurement of teamwork skills. *Journal of Workplace Learning*, 29(2), 124–133. [\[CrossRef\]](#)
- Camara, W., O'Connor, R., Mattern, K., & Hanson, M. A. (2015). Beyond academics: A holistic framework for enhancing education and workplace success. In *ACT research report series*. ACT, Inc. Available online: <https://eric.ed.gov/?id=ED558040> (accessed on 17 January 2026).
- Carter, E. W., & Lunsford, L. B. (2005). Meaningful work: Improving employment outcomes for transition-age youth with emotional and behavioral disorders. *Preventing School Failure: Alternative Education for Children and Youth*, 49(2), 63–69. [\[CrossRef\]](#)
- Chai, H., Hu, T., & Wu, L. (2024). Computer-based assessment of collaborative problem solving skills: A systematic review of empirical research. *Educational Research Review*, 43, 100591. [\[CrossRef\]](#)
- Cheung, G. W., Cooper-Thomas, H. D., Lau, R. S., & Wang, L. C. (2024). Reporting reliability, convergent and discriminant validity with structural equation modeling: A review and best-practice recommendations. *Asia Pacific Journal of Management*, 41(2), 745–783. [\[CrossRef\]](#)
- Chu, S. K. W., Reynolds, R. B., Tavares, N. J., Notari, M., & Lee, C. W. Y. (2021). *21st century skills development through inquiry-based learning from theory to practice*. Springer International Publishing. [\[CrossRef\]](#)
- Commission for the European Communities. (2008). *New skills for new jobs anticipating and matching labour market and skills needs*. Available online: <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2008:0868:FIN:EN:PDF> (accessed on 17 January 2026).

- Csapó, B. (2014). A szegedi iskolai longitudinális program [Longitudinal program in Szeged]. In J. Pál, & Z. Vajda (Eds.), *Szegedi Egyetemi Tudástár 7. Bölcsészeti- és társadalomtudományok* (pp. 117–166). Szegedi Egyetemi Kiadó.
- Csapó, B., Lőrincz, A., & Molnár, G. (2012). Innovative assessment technologies in educational games designed for young students. In D. Ifenthaler, D. Eseryel, & X. Ge (Eds.), *Assessment in game-based learning: Foundations, innovations, and perspectives* (pp. 235–254). Springer. [CrossRef]
- Csapó, B., & Molnár, G. (2019). Online diagnostic assessment in support of personalized teaching and learning: The eDia system. *Frontiers in Psychology*, 10, 1522. [CrossRef]
- Ferreira, C., Robertson, J., & Pitt, L. (2023). Business (un) usual: Critical skills for the next normal. *Thunderbird International Business Review*, 65(1), 39–47. [CrossRef]
- Greiff, S. (2012). From interactive to collaborative problem solving: Current issues in the Programme for International Student Assessment. *Review of Psychology*, 19(2), 111–121.
- Griffin, P., & Care, E. (Eds.). (2015). Assessment & teaching of 21st century skills. In *Methods and approach*. Springer. [CrossRef]
- Griffin, P., Care, E., & McGaw, B. (2012). *Assessment and teaching of 21st century*. Springer. [CrossRef]
- Hakkarainen, A. M., Holopainen, L. K., & Savolainen, H. K. (2016). The impact of learning difficulties and socioemotional and behavioural problems on transition to postsecondary education or work life in Finland: A five-year follow-up study. *European Journal of Special Needs Education*, 31(2), 171–186. [CrossRef]
- Han, A., Krieger, F., Borgonovi, F., & Greiff, S. (2023). Behavioral patterns in collaborative problem solving: A latent profile analysis based on response times and actions in PISA 2015. *Large-Scale Assessments in Education*, 11(1), 35. [CrossRef]
- Hao, J., Liu, L., von Davier, A. A., & Kyllonen, P. C. (2017). Initial steps towards a standardized assessment for collaborative problem solving (CPS): Practical challenges and strategies. In A. A. von Davier, M. Zhu, & P. C. Kyllonen (Eds.), *Innovative assessment of collaboration* (pp. 135–156). Springer. [CrossRef]
- Hastie, C., Fahy, K., & Parratt, J. (2014). The development of a rubric for peer assessment of individual teamwork skills in undergraduate midwifery students. *Women and Birth*, 27(3), 220–226. [CrossRef]
- Hesse, F., Care, E., Buder, J., Sassenberg, K., & Griffin, P. (2015). A framework for teachable collaborative problem solving skills. In P. Griffin, & E. Care (Eds.), *Assessment & teaching of 21st century skills. methods and approach* (pp. 37–56). Springer. [CrossRef]
- Józsa, K., & Morgan, G. A. (2017). Reversed items in Likert scales: Filtering out invalid responders. *Journal of Psychological and Educational Research*, 25(1), 7–25.
- Karaca-Atik, A., Meeuwisse, M., Gorgievski, M., & Smeets, G. (2023). Uncovering important 21st-century skills for sustainable career development of social sciences graduates: A systematic review. *Educational Research Review*, 39, 100528. [CrossRef]
- Kozma, R. (2009). Assessing and teaching 21st century skills: A call to action. In F. Schueremann, & J. Björnsson (Eds.), *The transition to computer-based assessment: New approaches to skills assessment and implications for large scale assessment* (pp. 13–23). European Communities.
- Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: A literature review. *Quality & Quantity*, 47(4), 2025–2047. [CrossRef]
- Lewis, K. (2003). Measuring transactive memory systems in the field: Scale development and validation. *Journal of Applied Psychology*, 88(4), 587–604. [CrossRef]
- Liu, L., Hao, J., von Davier, A. A., Kyllonen, P., & Zapata-Rivera, D. (2016). A tough nut to crack: Measuring collaborative problem solving. In Y. Rosen, S. Ferrara, & M. Mosharraf (Eds.), *Handbook of research on technology tools for real-life skill development* (pp. 344–359). IGI Global. [CrossRef]
- Loughry, M. L., Ohland, M. W., & Moore, D. (2007). Development of a theory-based assessment of team member effectiveness. *Educational and Psychological Measurement*, 67(3), 505–524. [CrossRef]
- Masters, G. N. (1982). A rasch model for partial credit scoring. *Psychometrika*, 47(2), 149–174. [CrossRef]
- Molnár, G., & Csapó, B. (2019). Making the psychological dimension of learning visible: Using technology-based assessment to monitor students' cognitive development. *Frontiers in Psychology*, 10, 1368. [CrossRef]
- Ng, Z. J., Lin, S., Niu, L., & Cipriano, C. (2025). Measurement invariance of the children's social desirability scale–short version (CSD-S) across gender, grade level, and race/ethnicity. *Assessment*, 32(3), 394–404. [CrossRef] [PubMed]
- OECD. (2010). *The definition and selection of key competencies* [Executive Summary]. Available online: [https://one.oecd.org/document/EDU/EDPC/ECEC/RD\(2010\)26/en/pdf](https://one.oecd.org/document/EDU/EDPC/ECEC/RD(2010)26/en/pdf) (accessed on 17 January 2026).
- OECD. (2017). *PISA 2015 results (volume V): Collaborative problem solving*. OECD.
- Olsen, J. K., Aleven, V., & Rummel, N. (2017). Exploring dual eye tracking as a tool to assess collaboration. In A. A. von Davier, M. Zhu, & P. C. Kyllonen (Eds.), *Innovative assessment of collaboration* (pp. 157–172). Springer. [CrossRef]
- O'Neil, H. F., Chuang, S., & Chung, G. K. W. K. (2003). Issues in the computer-based assessment of collaborative problem solving. *Assessment in Education*, 10(3), 361–373. [CrossRef]
- Rios, J. A., Ling, G., Pugh, R., Becker, D., & Bacall, A. (2020). Identifying critical 21st-century skills for workplace success: A content analysis of job advertisements. *Educational Researcher*, 49(2), 80–89. [CrossRef]

- Rojas, M., Nussbaum, M., Chiuminatto, P., Guerrero, O., Greiff, S., Krieger, F., & van der Westhuizen, L. (2021). Assessing collaborative problem-solving skills among elementary school students. *Computers & Education*, 175, 104313. [\[CrossRef\]](#)
- Roshid, M. M., & Haider, M. Z. (2024). Teaching 21st-century skills in rural secondary schools: From theory to practice. *Heliyon*, 10(9), e30769. [\[CrossRef\]](#)
- Salas, E., Reyes, D. L., & Woods, A. L. (2017). The assessment of team performance: Observations and needs. In A. A. von Davier, M. Zhu, & P. C. Kyllonen (Eds.), *Innovative assessment of collaboration* (pp. 21–36). Springer. [\[CrossRef\]](#)
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, 17(4), 2. [\[CrossRef\]](#)
- Smith-Jentsch, K. A., Cannon-Bowers, J. A., Tannenbaum, S. I., & Salas, E. (2008). Guided team self-correction: Impacts on team mental models, processes, and effectiveness. *Small Group Research*, 39(3), 303–327. [\[CrossRef\]](#)
- Sortwell, A., Trimble, K., Ferraz, R., Geelan, D. R., Hine, G., Ramirez-Campillo, R., Carter-Thuiller, B., Gkintoni, E., & Xuan, Q. (2024). A systematic review of meta-analyses on the impact of formative assessment on K-12 students' learning: Toward sustainable quality education. *Sustainability*, 16(17), 7826. [\[CrossRef\]](#)
- Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2008). The developmental psychometrics of big five self-reports: Acquiescence, factor structure, coherence, and differentiation from ages 10 to 20. *Journal of Personality and Social Psychology*, 94(4), 718. [\[CrossRef\]](#)
- Stehle, S. M., & Peters-Burton, E. E. (2019). Developing student 21 st Century skills in selected exemplary inclusive STEM high schools. *International Journal of STEM Education*, 6, 39. [\[CrossRef\]](#)
- Stoeffler, K., Rosen, Y., Bolsinova, M., & von Davier, A. A. (2020). Gamified performance assessment of collaborative problem solving skills. *Computers in Human Behavior*, 104, 106036. [\[CrossRef\]](#)
- Strom, P. S., Strom, R. D., & Wang, C. H. (2024). Peer and self-assessment of teamwork skills in high school: Using a multi-rater evaluation method for cooperative learning groups. *International Journal of Educational Reform*, 33(1), 81–100. [\[CrossRef\]](#)
- Taggar, S., & Brown, T. C. (2001). Problem-solving team behaviors: Development and validation of BOS and a hierarchical factor structure. *Small Group Research*, 32(6), 698–726. [\[CrossRef\]](#)
- Thornhill-Miller, B., Camarda, A., Mercier, M., Burkhardt, J. M., Morisseau, T., Bourgeois-Bougrine, S., Vinchon, F., El Hayek, S., Augereau-Landais, M., Mourey, F., Feybesse, C., Sundquist, D., & Lubart, T. (2023). Creativity, critical thinking, communication, and collaboration: Assessment, certification, and promotion of 21st century skills for the future of work and education. *Journal of Intelligence*, 11(3), 54. [\[CrossRef\]](#)
- Tushar, H., & Sooraksa, N. (2023). Global employability skills in the 21st century workplace: A semi-systematic literature review. *Heliyon*, 9, e21023. [\[CrossRef\]](#) [\[PubMed\]](#)
- Von Davier, A. A., Zhu, M., & Kyllonen, P. C. (Eds.). (2017). *Innovative assessment of collaboration*. Springer. [\[CrossRef\]](#)
- Voogt, J., & Roblin, N. P. (2012). A comparative analysis of international frameworks for 21st century competences: Implications for national curriculum policies. *Journal of Curriculum Studies*, 44(3), 299–321. [\[CrossRef\]](#)
- Zhuang, X., MacCann, C., Wang, L., Liu, L., & Roberts, R. D. (2008). Development and validity evidence supporting a teamwork and collaboration assessment for high school students. *ETS Research Report Series*, 2008(2), 1–51. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.