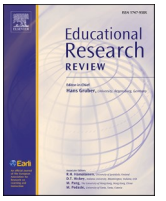
Educational Research Review 43 (2024) 100591

Contents lists available at ScienceDirect

Educational Research Review

journal homepage: www.elsevier.com/locate/edurev



Review

Computer-based assessment of collaborative problem solving skills: A systematic review of empirical research

Huanyou Chai [a,b,*], Tianhui Hu [a], Li Wu [c]

[a] Faculty of Artificial Intelligence in Education, Central China Normal University, Wuhan, 430079, Hubei, China
[b] Research Center of Distance Education, Beijing Normal University, Beijing, 100875, China
[c] School of Education, Central China Normal University, Wuhan, 430079, Hubei, China

ARTICLE INFO

Keywords:
Computer-based assessment
Collaborative problem solving skills
21st century skills
Theoretical model
Evaluation methodologies

ABSTRACT

Given the widespread concern on collaborative problem solving (CPS) skills, there has been an increasing interest in the last few years to explore how to assess them with digital technologies. This study systematically reviewed how CPS skills have been assessed with digital technologies in the literature. A total of 40 articles were reviewed to analyze specific computer-based assessment instruments of CPS skills from four perspectives: research context, theoretical model for assessment, assessment type, and reliability and validity evidence. The results indicate that most tests target a sample of less than 500 junior students. Nine theoretical models are employed for assessing CPS skills, most of which treat these skills as an explicit combination of social and cognitive skills and are applied to a limited range of participants' age levels, collaboration features, and team compositions. A total of 22 tests have been employed and fallen into four types, i.e., the ones with specific predefined messages in human-agent mode, and those with online chat box, videoconferencing, and face-to-face collaboration in human-human mode. Each type of these tests demonstrates great diversities in participants' age levels, types of CPS task(s), team compositions, types of assessment data, and methods of data recording and scoring. A certain number of tests lack reliability and validity evidence. Our findings are expected to benefit relevant researchers and test developers in terms of providing suggestions for future research which include testing the applicability of theoretical models for assessing CPS skills across a wide range of assessment contexts. In addition, future researchers should improve the development, data processing, and report of those four types of computer-based assessment instruments of CPS skills through different approaches, respectively.

1. Introduction

With the rapid development of modern society, an increasing number of projects or tasks require or benefit from teams of individuals with varying expertizes, backgrounds, and ideas to work in unison through communication and collaboration. For instance, learners often collaborate with their peers to achieve a shared understanding of course contents, and employees usually collaborate with their colleagues to solve complex problems at work (Dillenbourg & Traum, 2006; Rummel & Spada, 2005). In these or similar settings, be it formal or informal, team members are expected to be equipped with plenty of cognitive problem solving and social

* Corresponding author. Faculty of Artificial Intelligence in Education, Central China Normal University, Wuhan, 430079, Hubei, China.
E-mail addresses: chaihy@ccnu.edu.cn (H. Chai), huayuan2963291362@mails.ccnu.edu.cn (T. Hu), wuli1030@ccnu.edu.cn (L. Wu).

https://doi.org/10.1016/j.edurev.2023.100591
Received 9 June 2022; Received in revised form 16 November 2023; Accepted 23 December 2023
Available online 30 December 2023
1747-938X/

collaboration skills, i.e. collaborative problem solving (CPS) skills. From the cognitive perspective, a successful team must consist of members who precisely define the problem and identify the gap. From the social standpoint, team members must establish a shared understanding of the problem and take joint actions. Nowadays, CPS skills have drawn growing attention as an essential feature in the 21st century skills (Griffin, McGaw, & Care, 2012) and for their important role in the 21st century workforce (Burrus, Jackson, Xi, & Steinberg, 2013).

Against this background, CPS skills have gained increasing attention from researchers from diverse fields, including psychology, education, and sociology. Their work embraces, but is not confined to the following ones: (a) developing and validating theoretical models for assessing CPS skills (Hesse, Care, Buder, Sassenberg, & Griffin, 2015; Sun et al., 2020); (b) creating and iterating the assessment instruments of CPS skills (Rojas et al., 2021; Stadler, Herborn, Mustafić, & Greiff, 2020); (c) examining what significantly predicts CPS skills (Camacho-Morles et al., 2019; Tang, Liu, & Wen, 2021); (d) designing online courses to train or cultivate CPS skills (Rosen, Wolf, & Stoeffler, 2020; Song & Lan, 2018). These studies contribute to a large body of literature that provides valuable insights into the nature, assessments, and cultivation of CPS skills.

Of particular interest is the assessment of CPS skills, which has provided improved grounds for CPS training and applications. Traditionally, self-report questionnaires have been widely adopted to measure CPS skills (e.g., Fuad, Alfin, Fauzan, Astutik, & Prahani, 2019; Gu, Chen, Zhu, & Lin, 2015). They are characterized by capturing participants' self-confidence or self-efficacy for completing CPS tasks as a rough proxy for actual CPS skills. However, they are often criticized for being prone to various biases, such as social desirability bias and false memories (Gonyea, 2005; Holstein & Gubrium, 2001). In recent years, computer-based assessment (CBA) has gained great attention for allowing participants to directly demonstrate their CPS skills when engaging in a series of authentic or virtual CPS tasks (Aqlan & Zhao, 2022; Azura et al., 2021a). It is a powerful approach for evaluating CPS skills for its ability to concurrently provide both process and outcome data. Given a wide range of CBA instruments of CPS skills, it necessitates research into synthesizing the existing knowledge pertaining to this topic. This work contributes to achieving a comprehensive understanding of previous findings and guiding future research on developing new and improved assessments of CPS skills.

To date, few research has reviewed and synthesized the results of previous studies on assessing CPS skills. Oliveri, Lawless, and Molloy (2017) outlined the conceptualization, assessment, and validity considerations related to CPS skills and other related constructs, e.g. problem solving and teamwork. This review was limited only to empirical research published before 2015. Putri and Sinaga (2021) summarized the implementation, measurement, and development of CPS skills in science teaching and learning. What their work added to existing literature was a mere classification of the assessment environment of CPS skills into computer environment and real-life context. Baligar et al. (2020) reviewed the social and cognitive outcomes when assessing CPS skills in engineering education. To conclude, a systematic review is lacking on the state-of-the-art assessment instruments specifically focusing on CBA of CPS skills. Many researchers have point out that it is a challenging undertaking to accurately assessing CPS skills with CBA (Andrews-Todd and Forsyth, 2020; von Davier, Hao, Liu, & Kyllonen, 2017). Therefore, it necessitates systematic reflections to provide a clear picture of extant CBA instruments of CPS skills.

The present study is intended to systematically review empirical research on assessing CPS skills with CBA. It focuses on the CPS studies that develop or adopt specific CBA instrument to measure CPS skills in various settings. Specifically, we examine what research contexts to which assessments of CPS skills have been applied, what theoretical models existing studies have employed and what assessment contexts to which they have been applied, what types of assessments have been utilized and what assessment characteristics they have, and what the reliability and validity evidence are provided. As such, the present study aims to gain a full view of existing CBA instruments of CPS skills and to suggest future ways to guide development and implementation of CPS skills assessment.

1.1. Definition and significance of CPS skills

As a relatively new construct, CPS skills are psychologically defined as the set of skills of solving problems and working toward a common goal in collaboration with others in a team or group (O'Neil, Chuang, & Baker, 2010). Specifically, two key components of problem solving and social collaboration constitute the construct of CPS skills. Problem solving involves the abilities to transform a given state into a goal situation with cognitive efforts (Mayer & Wittrock, 2006), while social collaboration refers to the skills that allows a person to construct and share conception of a problem (Roschelle & Teasley, 1995). With the shift of workplace requirements, there is an increasing demand for students' and employee's proficiency in CPS skills for the last decades (Autor, Levy, & Murnane, 2003). In order to equip students and future career entrants with sufficient CPS skills, some intergovernmental economic organizations (e.g., Organization for Economic Co-operation and Development, OECD) and leading educational advocacy organizations (e.g. Partnership for 21st Century Learning) have emphasized the value of CPS skills and explored how to assess and cultivate them. Since then, CPS skills have gained widespread concern from educational researchers and practitioners as a core component in the 21st century skills (Griffin et al., 2012).

Combining problem solving and social collaboration skills together, the Program for International Student Assessment (PISA) proposed a new definition of CPS skills as "the capacity of an individual to effectively engage in a process whereby two or more computer agents attempt to solve a problem by sharing the understanding and effort required to come to a solution and pooling their knowledge, skills and efforts to reach that solution" (OECD, 2013, 2017). For the Assessment and Teaching of 21st Century skills (ATC21s) project, CPS skills are defined as "the abilities to recognize the point of view of other persons in a team; contribute knowledge, experience, and expertise in a constructive way; identify the need for contributions and how to manage them; recognize structure and procedure involved in resolving a problem; and as a member of the team, build and develop group knowledge and understanding" (Griffin et al., 2012). Notably, these two definitions are quite distinct from each other. The former focuses on individuals' collaboration with computer agents, while the latter concerns the collaboration between (among) human participants.

Moreover, there is no consensus on the finer-grained definitions or theoretical models for assessing CPS skills. Some scholars considered CPS skills as a one-dimension construct, e.g. Taylor and Baek (2018) adopted a survey to assess students' CPS skills from a single dimension perspective. Besides, many researchers separated the cognitive (or problem solving) aspects from the social (collaboration) ones of CPS skills in their proposed theoretical models. For example, the PISA framework combines three main collaborative competences with four stages of individual problem-solving process, resulting in a matrix of twelve specific skills (OECD, 2017). Similarly, ATC21S developed a theoretical framework comprised of three sets of social skills and two sets of cognitive skills. There are also some researchers integrating the cognitive and social aspects of CPS skills, as to a certain extent, these two aspects are contingent on each other (Care & Griffion, 2014). For example, Sun et al. (2020) proposed a generalized competency model of CPS skills consisting of three main facets with two sub-facets for each.

In addition, recent work has attempted to incorporate cultivating CPS skills into various educational practices. For example, Rosen et al. (2020) designed Animalia online mini-course to foster students' CPS skills in the context of complex ecosystems. Ostrander et al. (2020) developed an Intelligent Team Tutoring System for training pairs of learners to work collaboratively in a surveillance task. Another representative example is the work by Lin, Yu, Hsiao, Chang, and Chien (2020), who compared the effectiveness of web-based CPS systems and classroom-based collaborative hands-on learning activities in developing junior high school students' CPS skills.

Taken together, recent research has established the importance of teaching and cultivating CPS skills. However, as noted by Sun et al. (2020) and Stadler et al. (2020), there is a paucity of scientific justifications for the current definition and conceptualization of CPS skills. Unsurprisingly, various theoretical models would occur when researchers adopted, interpreted, and assessed the concept and definition of CPS skills.

1.2. Assessment of CPS skills

As a basis of systematic cultivation, assessment has long been recognized as an important driver to advance education on CPS skills since its inception (Griffin et al., 2012; OECD, 2013). Graesser et al. (2018) also suggested that much work should focus on improving the quality of assessing students' proficiency in CPS skills.

Numerous tests developed for gauging CPS skills are traditionally based on self-report questionnaires, in which participants are asked to evaluate how good they are at collaboratively solving certain problems. For example, Fuad, Alfin, FauzanAstutik, and Prahani (2019) employed five items to examine the effectiveness of group science learning to improve CPS skills of primary school teacher candidates. Taylor and Baek (2018) also used a survey to determine the positive effect of collaborative interventions on CPS skills for students working on collaborative robotics projects. According to Gonyea (2005) and Holstein and Gubrium (2001), this approach could only provide rough proxies for participants' actual skills. In addition, low correlations often exist in the relation between participants' self-reported and actual levels of certain skills (e.g., González-Betancor, Bolívar-Cruz, & Verano-Tacoronte, 2019; Hodes & Thomas, 2020). Hence, self-report questionnaires may not be an appropriate approach for capturing participants' CPS skills.

With the proliferation of computer-based technologies, CBA (often used interchangeably with computer-based testing, computer assisted assessment, and technology enhanced assessment, see Timmis, Broadfoot, Sutherland, & Oldfield, 2016) has been introduced into educational systems. They range from a simple delivery of paper-pencil tests as computerized versions, through to innovatively presenting tests with a video game or hypermedia (Perry, Meissel, & Hill, 2022). As indicated by Scherer, Greiff, and Kirschner (2017), CBA could create more innovative and perhaps more authentic item formats. Schacter, Herl, Chung, Dennis, and O'Neil (1999) argued that CBA is "a solution to the narrow measurement and reporting of (collaborative) problem-solving". To date, CBA has been adopted across various domains and contexts, especially in the field of assessing CPS skills.

Generally, CBA of CPS skills is characterized by allowing participants to verbally or nonverbally demonstrate their skills when executing one or more pre-designed CPS tasks in an authentic or virtual collaborative team (Bland & Gareis, 2018; O'Leary, Scully, Karakolidis, & Pitsia, 2018). According to the modes of assessment environments (Putri & Sinaga, 2021), CBA instruments of CPS skills could be grouped into online and face-to-face categories. For the former category, it requires participants to collaborate with one or more computer agents (human-to-agent (H-A) mode) or real humans (human-to-human (H–H) mode) in an online environment. For example, OECD applied the H-A mode to conduct large-scale standardized assessment of students' CPS skills across many countries (Tang et al., 2021; Yavuz & Atar, 2020). Yuan, Xiao, and Liu (2019) developed an online test in H–H mode to measure CPS skills with a new paradigm for extracting indicators and modeling the dyad data. The latter category requires participants to work with their peers (i.e., H–H mode) to collaboratively accomplish computer-based task(s) in a face-to-face environment. For example, Sun et al. (2020) assessed CPS skills in a team of three students sitting in a row to collaboratively play an educational game. Meanwhile, some research efforts have investigated the quality evidence of reliability and validity of CBA instruments of CPS skills. For instance, Stadler et al. (2020) tested the validity of *PISA 2015 CPS* tasks by identifying the relations between the assessment and existing collaboration measures. Rojas et al. (2021) conducted confirmatory factor analysis to explore the validity of their proposed instrument with two equivalent forms.

Notwithstanding plentiful tests employed in relevant studies, no study provides an in-depth and systematic insight into the characteristics of these assessments. Little is known about the research contexts where CPS skills have been evaluated, theoretical models of CPS skills that have been applied, assessment types that are employed, as well as the psychometric qualities of these tests. Without an integrated and clear view on these information, we will feel it beyond our capacity to improve and design effective tests for future requirements.

Therefore, we carried out a systematic review in this study in order to synthesize existing studies and to determine what needs to be explored by specifically focusing on the literature regarding CBA of CPS skills. We summarized the current state and characteristics of relevant studies and provided suggestions for future directions concerning how to inform the development and design of CBA of CPS

skills. More specifically, the research questions (RQs) were proposed as follows.

RQ1. What are the research contexts (e.g. participants' age levels and sample sizes) in which CPS skills have been assessed?

RQ2. What are the theoretical models that have been adopted for assessing CPS skills and what assessment contexts to which they have been applied?

RQ3. What types of assessments of CPS skills exist and what characterizes them (e.g., types of CPS task(s) in the assessment, team compositions, and types of assessment data, etc.)?

RQ4. What is the reliability and validity evidence gained from these assessments?

2. Method

To systematically collate empirical evidence on CBA of CPS skills, we followed the general procedures congruent with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Moher, Liberati, Tetzlaff, Altman, & The PRISMA Group, 2009). Fig. 1 demonstrates the detailed procedures in this review. We will detail the literature search procedure, inclusion/exclusion criteria, and literature coding procedure in the remainder of this section.

2.1. Literature search

The literature search was conducted in two widely used and comprehensive electronic databases to capture all CBA instruments used in studies on CPS skills: Web of Science (WoS) and Scopus. According to Chadegani et al. (2013), these two databases were the main ones for international multidisciplinary academic literature. We first gathered all articles with the only key phrase "collaborative problem solving" in all sections of each article. Given that some articles might contribute to assessment but do not specifically label it as a key term in their description, we did not add "assessment" as the search term in the procedure of literature search. Besides, although CPS is related to collaborative learning, problem solving, and teamwork, ample research has claimed that these concepts are essentially distinct from CPS (Graesser et al., 2018; Slavin, 2017). Therefore, we did no use these phrases as the search terms. After this step, a total of 1937 papers were identified electronically and duplicates (n = 214) were removed.

2.2. Inclusion/exclusion criteria

We adopted the following six inclusion criteria to select articles: (a) including "collaborative problem solving" in any section of the article (such as title, abstract, keywords, or main text); (b) published between 2011 and 2021; (c) available in full-text; (d) empirical studies containing the assessment of CPS skills; (e) involving the use of computer-based technologies in the assessment; and (f) written

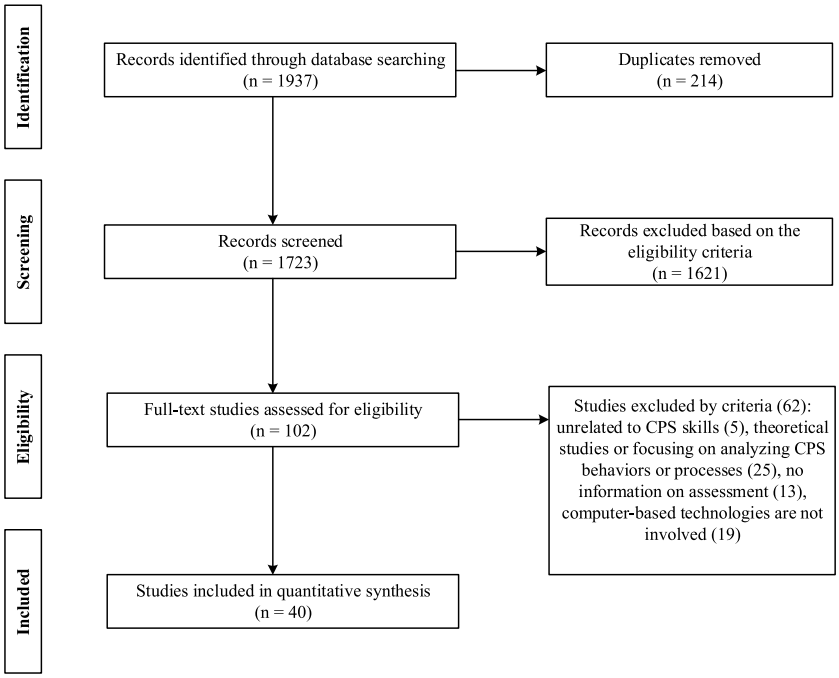


Fig. 1. Selection procedure flow chart.

in English. The exclusion criteria include: a) "collaborative problem solving skills" is not investigated by the research; (b) the article is a conceptual or theoretical work, or an analysis of CPS behaviors or processes, or a review of existing studies; (c) no information on assessment is reported; and (d) computer-based technologies are not used in the assessment.

After receiving training sessions from experts in systematic review, the three authors independently scanned the title, abstract, and keywords of each paper, and applied the above criteria to screen papers. This step made the number of the collected papers sharply decline to 102. Then the researchers read the paper in full-text and applied the inclusion/exclusion criteria. By doing so, we finally selected 40 articles as relevant for data extraction. Notably, disagreements among the authors were finally resolved through in-depth discussion and further examination of the controversial studies.

2.3. Literature coding

Drawing on the procedures of a content analysis (Fraenkel, Wallen, & Hyun, 2015), we coded the papers by systematically clarifying the texts into various categories in three stages. First, a coding scheme was created to systematically capture the information from sample papers, which echoed the four research questions. An Excel spreadsheet was specifically set up to store and analyze all data. Second, the three authors took three phases to refine the coding scheme. In phase 1, 12 papers were randomly chosen from the sample papers. In phase 2, we coded each of the first five papers by discussing collaboratively and further updated our coding scheme. In phase 3, each author first coded each of the remaining seven papers independently, and then we modified the coding scheme based on comparing and discussing the coding results, especially those with differences in categories. Third, after the coding scheme was optimized and the coders were well trained, the first and second authors further coded 10 papers independently. The inter-rater agreement is 0.92 across all the categories. For the discrepancies, the third author would be invited to engage in the discussions until all disagreements were eliminated. Fourth, given the acceptable inter-rater agreement, the first author coded the remaining papers independently. Appendix A shows the basic coding results of the sample papers included in this study.

3. Results

3.1. Research contexts of the CPS skills assessments

According to Appendix A, there were 10 countries/regions that developed or applied CBA of CPS skills; thereinto, OECD and ATC21S, two famous international organizations, developed one widely used assessment respectively. In addition, the tests in most studies ($n = 38$) were adopted to collect data in a single country/region; only two studies used the assessment of *PISA 2015 CPS* to test participants from two or more countries (Ham & Hwang, 2021; Tekin & Aktan, 2021). In particular, Tekin and Aktan (2021) conduct a cross-country comparison of measurement invariance of CPS skills with participants from Singapore, Turkey, and Norway. Ham and Hwang (2021) examined the relationship between mathematics achievement and CPS skills by analyzing the PISA 2015 data across two countries (USA and Korean).

Besides, the studies were published between 2015 and 2021. Particularly, there was an obvious growth of the number of studies from 2020 (see Fig. 2). Regarding participants (see Table 1), it ranged from 15 (Pöysä-Tarhonen et al., 2021) to 53855 participants (Kuo et al., 2020) in the sample size, covering elementary student up to older adult (age = 68). As shown in Table 1, the number of 100–500 participants was the most researched sample size that covered more than one-third of the reviewed studies, followed by the number of less than 100 participants (29.27%). The rest of the studies were conducted in a sample of 500–1000 (17.07%) or more than 1000 (17.07%) participants. Moreover, junior student was the age level most often researched (48.89%). A few studies targeted elementary student (13.33%), senior student.

(11.11%), undergraduate (11.11%), and adult (11.11%), respectively. There were also two studies (4.88%) that targeted a wide range of participants' age levels (age = (16–60)/(18–68)). To conclude, most studies recruited a sample of less than 500 participants composed of junior students.

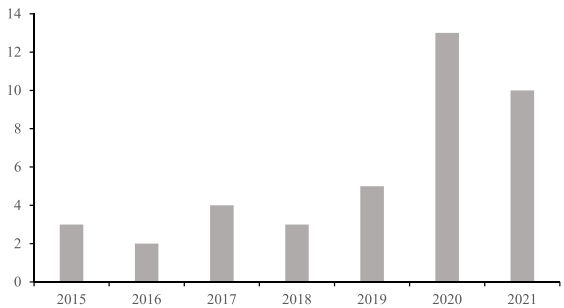


Fig. 2. Articles published by Year.

Table 1
Participants' sample size and age level of the reviewed studies.

Variables	Categories	Numbers	Percent
Sample size	<100	12	29.27%
	100–500	15	36.59%
	500–1000	7	17.07%
	>1000	7	17.07%
Age level	EL	6	13.33%
	JU	22	48.89%
	SE	5	11.11%
	UN	5	11.11%
	AD	5	11.11%
	Age = (16–60)/(18–68)	2	4.88%

Note: EL = Elementary school students, JU = Junior students, SE = Senior students, UN = Undergraduate, AD = Adults. Some studies adopted more than one categories so the total number (i.e., the denominator) to calculate the percentage is the total number of each category used in these studies.

3.2. Theoretical models for assessing CPS skills and the assessment contexts to which they have been applied

As noted in the introduction, great diverseness emerged in existing definition and operationalization of CPS skills when they were assessed. According to how the social and cognitive aspects of CPS skills are treated, we divided the total of nine theoretical models in the reviewed studies into the categories of separated and integrated model. Columns 1–4 in Table 2 shows the theoretical model(s) in each category and their proportions.

The separated model: More than 90 percent of the reviewed studies assessed CPS skills with the theoretical model separating the cognitive and social aspects (see Appendix B for an overview). This type of models shared the understanding that CPS skills were a combination of a set of cognitive and social skills. For instance, the model proposed by PISA (OECD, 2017) has received the most attention (42.5%, 17 studies). It divided CPS skills into four individual problem solving processes crossed with three social collaboration dimensions. Another prominent example was derived from ATC21S (Griffin et al., 2012), who classified CPS skills into three strands of social skills and two strands of cognitive skills. It ranked second (22.5%, 9 studies) in our proposed classification approach. Besides, three studies (7.5%) adopted the model proposed by Liu, Von Davier, Hao, Kyllonen, and Zapata-Rivera (2016), which documented a matrix composed of two cognitive skills and four social skills. ACT (American College Testing) created the Holistic Framework of CPS (5%, 2 studies) to support a more holistic understanding of the knowledge skills and behaviors required for success in college and career (Camara, O'Connor, Mattern, & Hanson, 2015). In this framework, CPS skills were divided into two major categories: Team effectiveness (similar to the set of social skills) and task effectiveness (similar to the set of cognitive skills). Andrews-Todd and Kerr (2019) devised an ontology-based model (5%, 2 studies) to assess CPS skills, in which five cognitive and four social skills were involved. Moreover, there were three studies (10%) that addressed CPS skills with other three different theoretical models. For example, the model of co-measure, proposed by Herro, Quigley, Andrews, and Delacruz (2017), was considered as a

Table 2
A classification of theoretical models for assessing CPS skills and the assessment contexts to which they have been applied.

Categories	Theoretical model	Frequency	Percent	Participants' age levels	Collaboration features	Team compositions
SEP (37, 92.5%)	The model by PISA	17	42.5%	EL (2), JU (12), SE (3), UN (1), AD (1)	H-A mode: SPMs (14); H–H mode: OCB(2), FAC (1)	1H/(1–3)A(s) (8), 1H/(1–2)A(s) (2), 1H/1A (2), 1H/2A (1), 1H/3A (1); 2Hs (1), 6Hs (1), (3–4)Hs (1)
	The model by ATC21S	9	22.5%	EL (3), JU (6), SE (1), AD (2)	H–H mode: OCB(8), FAC (1)	2Hs (7), (3–4)Hs (1), (3–5)Hs (2)
	Liu et al.'s (2016) model	3	7.5%	AD (2), Age = 18–68 (1)	H–H mode: OCB(3)	2Hs (3)
	The Holistic framework	3	5%	EL (1), JU (2), AD (1)	H-A mode: SPMs (1)	1H/1A (2), No info (1)
	The model from an ontology	2	5%	UN (1), Age = 16–60 (1)	H–H mode: OCB(3)	(3–5)Hs (1), 3Hs (1)
	Co-measure	1	2.5%	EL (1)	H–H mode: FAC (1)	Dynamic (1)
	Azura et al.'s model (2021)	1	2.5%	SE (1)	H–H mode: OCB(1)	3Hs (1)
	Krkovic et al.'s model (2016)	1	2.5%	JU (1)	H-A mode: SPMs (1)	1H/1A (1)
INT (3, 7.5%)	The generalized competence model	3	7.5%	JU (1), UN (3)	H–H mode: VIF (3), FAC (1)	3Hs (4)

Notes: SEP and INT represent that the separation and integration of the social and cognitive aspects of CPS skills in the model, respectively.
SPMs = Specific predefined messages, OCB= Online chat box, VIF = videoconferencing, FAC = Face-to-face collaboration; H= Human, A = Agent.
Some studies had more than one kind of assessment characteristic categories so the total number was more than 40.

validated rubric for assessing students' CPS skills in makerspace activities (Herro, Quigley, & Abimbade, 2021). It conceptualized CPS skills as a combination of two social (positive communication and peer interaction) and two cognitive (inquiry-rich/multiple paths and transdisciplinary learning) dimensions. In Azura et al.(2021)a study, students' CPS skills were divided into five aspects: participation skills, perspective-taking skills, and social regulation in the social domain, task regulation and knowledge building in the cognitive domain. Another theoretical model was from Krkovic et al. (2016), who measured CPS skills as a construct of two problem-solving dimensions and three collaboration dimensions.

The integrated model: In this type of theoretical models, the cognitive and social aspects were explicitly integrated with each other when CPS skills were assessed. Only the generalized model of CPS (Sun et al., 2020) was identified (7.5%, 3 studies). As described in the above section, most theoretical models treat the social and cognitive aspects as separate parts with their own processes. Instead, the generalized model of CPS skills is built on the premise that these two aspects are contingent on each other (Care & Griffion, 2014). According to Sun et al. (2020), CPS skills could be summarized as a combination of three main facts at the first level in a hierarchical structure. They could further be classified into multiple categories of sub-facets at the second level and associated behavioral indicators at the third level.

To acquire an in-depth understanding of the theoretical models for assessing CPS skills, Columns 5–7 in Table 2 presents the categorizations of the reviewed studies regarding the assessment contexts to which these models have been applied. The number in the parentheses at the end of each category of the assessment context characteristics indicates the number of studies that shows the corresponding characteristic. Specifically, the categorizations, based on a comprehensive review of the methodology characteristics of each study, were divided respectively by: (a) type of assessment participants' age levels; (b) type of collaboration features; (c) type of team compositions. Specifically, *collaboration feature* is used to indicate whether participants are assessed in H-A or H–H mode and the way in which they collaborative with other members during the assessment. It includes four categories of specific predefined messages in H-A mode, and online chat box, videoconferencing, and face-to-face collaboration in H–H mode. For the first category, participants are required to collaborate with one or more computer agents by exchanging predefined messages in a chat box. For the remaining three categories, participants could freely collaborate with one or more real humans in an online chat box, videoconferencing, or face-to-face collaboration, respectively. *Team composition* is adopted to indicate what constituted the collaborative team for executing the CPS task(s) in the assessment. It could be categorized by the number of agents or human participants involved in the assessment. For example, the '1H/(1–3)A(s)' represents that the CPS team consists of a human being assessed and a minimum of one and maximum of three agents in various CPS tasks in the assessment, while '(3–5)Hs' represents that the participant number in the assessment ranges from three to five humans in various CPS tasks. One exceptional category is: 'Dynamic', indicating that the number of the team members is not fixed and may change across time during the CPS task(s).

Regarding participants' age levels, the model by PISA was used across a wide range of age levels from elementary to university student, and junior student was the most researched age level. To be specific, twelve studies (70.58%, e.g., Ham & Hwang, 2021; Herborn, Stadler, Mustafić, & Greiff, 2020) adopting the model by PISA targeted junior students. Similarly, studies adopting the model by ATC21S presented a wide range of age levels except for undergraduate. They were also mainly focused on junior students (6 studies, 66.67%, e.g., Ahonen & Harding, 2018; Harding, Griffin, Awwal, Alom, & Scoular, 2017). Besides, Liu et al.'s (2016) model was only used for participants whose age was above 18 (e.g., Hao et al., 2015; Hao, Liu, von Davier, Kyllonen, & Kitchen, 2016), while the model from an ontology was only for participants above 16 years old (Andrews-Todd and Forsyth, 2020; Lin, Dowell, & Godfrey, 2021). Furthermore, the Holistic framework was used for participants at levels of elementary and junior student and adult (e.g., Polyak, von Davier, & Peterschmidt, 2017; Stoeffler et al., 2020), the generalized competence model was for junior and university student (e.g., Stewart, Keirn, & D'Mello, 2021; Sun et al., 2020), while the remaining models were only for participants at a single age level (e.g., Azura et al., 2021a; Krkovic et al., 2016).

Concerning the collaboration features, the model by PISA was used for tests with all four types of collaboration features. Specifically, this model was most commonly found for tests with specific predefined messages in H-A mode (14 studies, 82.35%, e.g., Ham & Hwang, 2021; Herborn et al., 2020). Besides, the model by ATC21S was mainly used for tests with online chat box in H–H mode (8 studies, 88.89%, e.g., Ahonen & Harding, 2018; Harding et al., 2017), while the generalized competence model was mainly for tests with videoconferencing in H–H mode (3 studies, 75%, e.g., Stewart et al., 2021; Sun et al., 2020). Furthermore, the remaining models were only used for tests with a single type of collaboration features. For example, Liu et al.'s (2016) model and the model from an ontology were only used for tests with online chat box in H–H mode (Hao et al., 2015, 2016).

As to the team compositions, the model by PISA was used in a wide range of the numbers and types of team members. Among related 17 studies, 14 studies (82.35%) adopted the team composition of one human being assessed and various numbers of agents ranging from one to three in various CPS tasks (e.g., Ham & Hwang, 2021; Herborn et al., 2020). Another three studies adopted the team composition of two (Nouri, Åkerfeldt, Fors, & Selander, 2017), three to four (Song, Park, & Park, 2020), and six (Song & Lan, 2018) humans, respectively. Besides, the model by ATC21S and that from an ontology were respectively used for two different types of team compositions (2Hs, (3–4)Hs; (3–5)Hs, 3Hs). Specifically, the former model was mostly used in a team of two humans (8 studies, 88.89%, e.g., Ahonen & Harding, 2018; Camacho-Morles et al., 2019). Furthermore, the remaining models were all used for a single type of team compositions (e.g., Azura et al., 2021a; Krkovic et al., 2016). Exceptionally, the team composition in the single study guided by the framework of Co-measure was dynamic in the number of human participants (Herro et al., 2021).

To conclude, the above review indicates that most studies adopted a theoretical model that treated CPS skills as an explicit combination of social and cognitive skills. In addition, most of these theoretical models were adopted in a limited range of participants' age levels, collaboration features, and team compositions. Finally, it is worth noting that most studies have not specified why certain theoretical model could be adopted for assessing CPS skills in the specific assessment context.

Table 3
A classification of tests of CPS skills and their characteristics.

Categories	Test or task name	Frequency	Participants' age level	Type of CPS task(s)	Team compositions	Type of Assessment data	Method of data recording	Method of data scoring	Reliability evidence	Validity evidence
H-A mode: SPMs (18)	*PISA 2015 CPS*	9	JU (8), SE (3), AD (1)	FRE	1H/(1–3)As	SPMs	Log file	Automatic	X	X
	Assessment System for CPS skills	2	JU (2)	FRE	1H/(1–2)As	SPMs	Log file	Automatic	–	X
H-A mode: SPMs (18)	*Circuit Runner*	2	JU (2)	FRE	1H/1A	SPMs	Log file	Automatic	X	X
	Zoo Quest	2	JU (2)	DEP	1H/1A	SPMs	Log file	Automatic	–	–
	A constellation or molecule building task	1	EL (1)	FRE	1H/2As	SPMs	Log file	Automatic	X	X
	Science Fair	1	EL (1). JU (1)	DEP	No info	SPMs	Log file	Automatic	–	–
	A self-development test (a)	1	JU (1)	FRE	1H/1A	SPMs	Log file	Automatic	–	X
H–H mode: OCB (17)	*ATC21S CPS*	7	EL (1), JU (3), SE (2), AD (1), age = 18–68 (1)	Both	2Hs	Actions, chats	Log file	Automatic	X	X
	A collaborative simulation task	2	AD (2)	DEP	2Hs	Actions, chats	Log file	Human	–	–
	Problems related to cyberbullying learning and game design	2	EL (2), JU (1)	FRE	(3–5)Hs	Actions, chats	Log file	Automatic	–	–
	A physics CPS task	1	SE (1)	DEP	3Hs	Actions, chats	Log file	Automatic	X	–
	A balancing scale problem task	1	JU (1)	DEP	2Hs	Actions, chats	Log file	Human	–	X
	Concept Map	1	UN (1)	DEP	(3–4)Hs	Actions, chats	Log file	Human	–	–
	Collaborative Science Assessment Prototype	1	UN (1)	DEP	(3–5)Hs	Actions, chats	Log file	Human	–	–
	Three-Resistor Activity	1	Age = 16-60	DEP	3Hs	Actions, chats	Log file	Human	X	X
H–H mode: OCB (17)	A self-development test (b)	1	JU (1), SE (1)	DEP	2Hs	Actions, chats	Log file	Automatic	–	X
	A self-development test (c)	1	JU (1)	FRE	2Hs	Actions, chats	Log file	Both	X	X
H–H mode: VIF (3)	*Minecraft Hour of Code*	3	UN (3)	DEP	3Hs	Actions, chats, Facial expressions	Video	Human (2), Both (1)	X	X
H–H mode: FAC (4)	A biology CPS task	1	EL (1)	DEP	6Hs	Chats, facial expressions, physical interactions	Video	Human	X	X
	Makerspace activities	1	EL (1)	FRE	Dynamic	Actions, chats, physical interactions	Video	Human	X	X
	Physics Playground	1	JU (1)	DEP	3Hs	Actions, chats, physical interactions	Video	Human	–	X
	Virtual manipulatives	1	AD (1)	DEP	(3–4)Hs	Actions, written reports, physical interactions	Video	Human	–	–

Note: Letters in the parentheses after test name or task were used to distinguish between different ones.

3.3. Assessments of CPS skills and their characteristics

In Table 3, a total of 22 tests (see Column 2) were adopted to assess CPS skills. Most of them (14 tests, 63.64%) were adopted in a single study, whereas five tests were respectively presented in two studies. There were also three tests, i.e., *Minecraft Hour of Code* (Stewart & D'Mello, 2018), *ATS21S CPS* (Griffin & Care, 2015; *PISA 2015 CPS* (OECD, 2017), that were respectively adopted in three, seven, and nine studies. Besides, seven tests assessed CPS skills with the model by PISA (OECD, 2017), while five with the model by ATC21S (Griffin et al., 2012). There was an equal number (i.e., two) of tests using the Holistic Framework, ontology-based model, and generalized model. The remaining four tests adopted four different models, respectively.

To capture a deep view of these tests, we classified them into four categories (see Column 1) according to the collaboration features in the assessment. Besides, we elaborated the assessment characteristics of each test in terms of participants' age levels, types of CPS task(s), team compositions, types of assessment data, and methods of data recording and scoring. Noticeably, each test may include one or more CPS tasks for assessing participants' CPS skills. According to whether subject-related knowledge and skills are needed for executing the CPS task(s), the assessment task(s) could be categorized into task(s) of content-free and -dependent categories.

In 18 studies (45%), CPS skills were assessed by seven tests with specific predefined messages in H-A mode (e.g., Ham & Hwang, 2021; Lin et al., 2020). These tests are based on the latest digital technologies that construct computer-simulated collaborative team consisting of one participant being assessed and varying number of computer agents to complete one or more pre-designed CPS task(s). In each task, participants could collaborate with the agent(s) by exchanging predefined messages in a chat box. Among the seven tests, *PISA 2015 CPS* was the most frequently used one (9 studies, e.g., Ham & Hwang, 2021; Herborn et al., 2020). Besides, there were seventeen studies (42.5%) that adopted ten different tests with online chat box in H–H mode (e.g., Azura et al., 2021a; Camacho--Morles et al., 2019). These tests draw on the state-of-the-art computer-based technologies to support two or more real humans to collaborate by an online chat box to complete one or more pre-defined task(s). When executing the CPS tasks, participants are free to type, say, or do whatever they want, even those that deviate from the CPS process. Among the ten tests, *ATC21S CPS* was the most commonly used test (7 studies, e.g., Camacho-Morles et al., 2019; Scoular & Care, 2020). Furthermore, three studies (7.5%) adopted the only test with videoconferencing in H–H mode (i.e., *Minecraft Hour of Code*) to assess CPS skills (e.g., Stewart et al., 2021; Sun et al., 2020). It was virtually the same as those with online chat box in H–H mode except in the collaboration medium of videoconferencing rather than online chat box. Lastly, there were another four studies (10%) that adopted four different tests with face-to-face collaboration in H–H mode, respectively (e.g., Herro et al., 2021; Sun et al., 2020). In this case, two or more humans are brought together to collaboratively complete one or more pre-designed CPS tasks in a face-to-face environment, assisted by some computer-based technologies such as Google Docs, Slides, Seesaw videos, and educational games. This type of assessments is quite different from the others in that only it allows for physical interactions when completing the pre-designed CPS task(s). For example, Sun et al. (2020) assessed CPS skills in a team of three students who collaboratively played an educational game in a face-to-face environment, in which one student controlled the mouse and the other two were asked to give additional assistance.

In the column of participants' age levels of Table 3, *ATC21S CPS* was adopted for the widest range of participants' age levels from elementary student to adult. Similarly, *PISA 2015 CPS* was also employed for a wide range except for the level of elementary student. Another test, *Three-Resistor Activity*, was used in a sample of participants whose age was from 16 to 60. However, the remaining tests (e.g. *Physics Playground* and *Minecraft Hour of Code*) were all adopted for a narrow range of participants' age levels. For example, *Minecraft Hour of Code* was only adopted for undergraduates (Sun et al., 2020).

Concerning the types of CPS task(s), a total of twenty-five studies (62.5%) adopted the tests involving content-free task(s), while thirty studies (75%) adopted the tests involving content-dependent task(s). In particular, seven studies (17.5%) adopted *ATC21S CPS*, which was a special case that involved both content-free and -dependent tasks (e.g., Camacho-Morles et al., 2019; Scoular & Care, 2020). Further examination of the contents in the assessments showed that only two studies were related with the course contents in language and literacy and information technology (e.g. computer programming), respectively, while the remaining studies were with sciences, including mathematics, physics, and biology. Besides, for the eighteen studies employing the tests with specific predefined messages in H-A mode, most of them (15 studies, 83.33%) adopted content-free tasks (e.g., Ham & Hwang, 2021; Lin et al., 2020). For the seventeen studies employing the test with online chat box in H–H mode, fifteen studies (88.24%) adopted content-dependent task (s) (e.g., Andrews-Todd and Forsyth, 2020; Harding et al., 2017), while ten studies (58.82%) adopted content-free task(s) (e.g., Scoular & Care, 2020; Yuan et al., 2019). And, three studies adopted the test with videoconferencing in H–H mode that only involved a content-dependent task (e.g., Stewart et al., 2021; Sun et al., 2020). Most of the studies (3 studies, 75%) employing the tests with face-to-face collaboration in H–H mode were concerned with content dependent task(s) (e.g., Song & Lan et al., 2018; Sun et al., 2020).

Regarding team compositions, they were quite different across various types of assessments. For tests with specific predefined messages in H-A mode, a collaborative team was usually composed of one participant being assessed and one to three computer agents. For example, in the assessment of *PISA 2015 CPS*, students were asked to collaborate with a minimum of one to a maximum of three agents to solve a realistic life problem in each task (e.g., Camacho-Morles et al., 2019; Scoular & Care, 2020). For tests with online chat box in H–H mode, the majority of studies (12 studies, 70.59%) reported a collaborative team of two real humans. For example, the assessment of *ATC21S CPS* asked participants to engage in dyadic communication supported by an online chat box (e.g., Andrews-Todd and Forsyth, 2020; Harding et al., 2017). For the test with videoconferencing in H–H mode, three real humans constituted a collaborative team during the assessment process. For example, in Sun et al.'s (2020) one study on assessing CPS skills, three students were connected together by video conferencing to jointly solve a computer programming task. For tests with face-to-face collaboration in H–H mode, the team composition was of three or more than three real humans. As exemplified above, one study conducted by Sun et al. (2020) adopted the team composition of three students in a face-to-face environment.

In terms of the types of assessment data, they showed great diversities in those four types of assessments. First, what tests with

specific predefined messages in H-A mode provided was numerous specific predefined messages, which were selected as their responses to assessment stimuli from the participants. As described above, *PISA 2015 CPS* supported participants to collaborate with specific predefined messages (OECD, 2017). Second, tests with online chat box in H–H mode could help produce the assessment data of online actions and chats, as exemplified by *ATC21S CPS* (e.g., Andrews-Todd and Forsyth, 2020; Scoular & Care, 2020). Notably, actions herein refer to participants' interactions with the computer-based technologies or tools adopted in the assessment. Third, participants assessed by the test with videoconferencing in H–H mode would offer the assessment data of facial expressions apart from actions and chats, as presented in the study of Sun et al. (2020). Fourth, participants assessed by the last type of assessments could provide a wide variety of assessment data, including actions, chats, facial expressions, physical interactions, and written reports (e.g., Song & Lan et al., 2018; Sun et al., 2020). Specifically, the unique affordance of physical interactions in this type of assessments made the assessment context quite close to authentic CPS environments.

As to the methods of data recording and scoring, for tests with specific predefined messages in H–H mode, the assessment data were recorded by the log files and analyzed automatically by the computer according to the predefined rubrics (e.g., Herborn et al., 2020; Rojas et al., 2021). Besides, all studies employing the tests with online chat box in H–H mode used log files to record the assessment data and the majority of them (9 studies, 52.94%) only used manual coding to score participants' performance (e.g., Nouri et al., 2017; Song et al., 2020). Rather, the assessment data of actions in *ATC21S CPS* and a self-developed test in the study of Scoular and Care (2020) was scored automatically by the pre-written algorithms (Adams et al., 2015). However, automatically scoring the chats was beyond the scope of existing studies. An exceptional study was from Yuan et al. (2019), who first scored the assessment data of chats of a few students by human coding and then applied their developed automatic scoring programs to automatically score the remaining data. In addition, researchers employing the test with videoconferencing in H–H mode videotaped participants' performance and usually manually coded the occurrences of behavioral indicators. Exceptionally, Stewart et al. (2021) followed the practice of Yuan et al. (2019) that manual coding was conducted prior to automatic scoring. Lastly, all four studies involving the tests with face-to-face collaboration in H–H mode adopted video devices to record participants' performance and manual coding to score the assessment data (e.g., Song & Lan et al., 2018; Sun et al., 2020).

3.4. Reliability and validity evidence of assessments of CPS skills

To address RQ4, we continue with the twenty-two tests in the reviewed studies: 10 were provided with reliability evidence and 14 validity evidence, and all the reported evidence was at an acceptable level. A total of 33 studies provided one or more indicators of reliability evidence, while 29 studies reported validity evidence. In particular, nine studies adopted the assessment from *PISA 2015 CPS* (OECD, 2017), whose reliability and validity results have been detailed in the study of Stadler et al. (2020). Seven studies adopted the assessment from *ATC21S CPS* (Griffin & Care, 2015), whose reliability and validity evidence could be found in the work of Griffin, Care, and Harding (2015).

Some reliability evidence (i.e., Cronbach's α, inter-rater consistency, or Cohen's kappa) was reported when human coding was implemented, especially for video or dialogue data from the tests in H–H mode (e.g., Song et al., 2020; Song & Lan, 2018). For instance, Song et al. (2020) asked two experts to rate participants in their CPS skills score and reported an acceptable level of Cohen's Kappa (0.95). Some researchers who drew on item response modeling usually employed the coefficient of expected a posteriori/plausible value (EAP/PV) reliability or separation reliability to calculate the reliability (e.g., Harding et al., 2017; Kuo et al., 2020). As reported by Kuo et al. (2020), the reliability was considered satisfactory with the EAP/PV coefficients of all dimensions of CPS skills being larger than 0.79.

Of the studies reporting validity, varied types of validity results were presented. For example, Nouri et al. (2017) investigated the predictive validity of a self-developed test by correlating students' score of CPS skills with their performance measures of accomplished tasks. In line with their hypotheses, the results showcased various degrees of significant correlations. Besides, in the study of Lin et al. (2015), three specialists were recruited to review the content validity of eight assessment modules in a web-based learning platform. Criterion-based validity was also examined by analyzing the Pearson product-moment inter-correlations of students' performance on the assessment and overall and sub-area skills. According to the data and analysis, these two forms of validity evidence were both judged to be good. In another case, Harding et al. (2017) compared students' CPS performance on pre-designed mathematical tasks with their performance on a set of content-free tasks to examine convergent and discriminant validity. As expected, students' scores on the social skills for the mathematical tasks were correlated with those on the social skills for the content-free tasks. By contrast, the correlations regarding students' scores on the cognitive skills varied according to the content area. To evaluate the construct validity of their newly-developed two versions of test, Rojas et al. (2021) employed exploratory factor analysis and confirmatory factor analysis with robust maximum-likelihood estimator. They found that the model achieved an excellent fit when all test items were grouped into 4 factors.

Although some quality information of reliability and validity evidence are reported for the assessments of CPS skills, we have to admit that a certain number of tests still lacked the quality information. Without this information, it is impossible for the stakeholders in educational field to accept and apply the corresponding tests with confidence. Thus, there is an increasing call for more reliability and validity evidence on extant tests of CPS skills to promote their large-scale application (e.g., Rojas et al., 2021; Stadler et al., 2020).

4. Discussion

Research on assessing CPS skills has stolen the limelight in academia and profoundly affected the practice of cultivating students' abilities in the last decades. Through this systematic review, we provided a comprehensive description of the current territory of CBA

of CPS skills regarding their related research contexts, theoretical models, assessment types, as well as quality evidence. To ensure this study's transparency and replicability, we strictly followed the procedures and methodology proposed by Moher et al. (2009).

4.1. Research contexts of the CPS skills assessments

According to our systematic review, CBA instruments of CPS skills were mainly developed by and applied in countries from USA, Europe, and East Asia. This finding, to a certain extent, indicates considerable international variation in how research is crossed with national initiatives to motivate the assessment of CPS skills. One possible reason may be that the economy in the above-mentioned countries is more global and industrialized, so that CPS skills are more indispensable in their labor market and more valued by their researchers and policy makers.

In addition, this review reported an obvious growth of research on the assessments of CPS skills from 2020, indicating the increasing emphasis on assessing these skills for their important role in the 21st century life, learning, and workplace. This finding is congruent with the wide application of OECD's (2017) 2015 PISA assessment and ATC21S assessment (Care, Griffin, & Wilson, 2018). Motivated by these two assessment programs, a large number of international organizations and scholars have been devoted to exploring how to make an accurate assessment of CPS skills. Concerning the research participants, this review revealed that the majority of tests focused on junior students. As evidenced in the literature, many key abilities related to CPS skills, such as abstract thinking (Dumontheil, 2014; Molnár, Greiff, & Csapó, 2013) and peer communication skills (Shaffer & Kipp, 2009), have gradually emerged or developed rapidly in the age of middle school. It is therefore not surprising to find many tests target participants at this age stage. However, to a certain extent, this finding reflects a relative lack of tests for measuring elementary and senior students and even adults. Therefore, more research is encouraged to develop tests for participants at these three age stages. On the other hand, the sample sizes of participants in the majority of studies were in small to moderate scale, indicating the need to include more participants in future studies to improve the trustworthiness of their findings.

4.2. Theoretical models for assessing CPS skills and the contexts to which they have been applied

Our review showed that a total of nine theoretical models were adopted for assessing CPS skills with eight in the separated type and one in the integrated type. Most of the selected studies adopted the separated model to define and conceptualize CPS skills as a combination of a set of social and cognitive skills. Among the eight models in this type, the ones from PISA and ATC21S were used more frequently by CPS researchers. These findings are in line with the widespread application and promotion of several international large-scale assessments involving CPS skills in education, e.g. OECD's (2017) 2015 PISA assessment and ATC21S assessment (Care et al., 2018). All these tests adopt the models that treat the cognitive and social aspects of CPS skills as separate dimensions. In contrast, only the generalized competence model of CPS skills was identified in the type of integrated model. It concerns the explicit integration of the cognitive and social aspects of CPS skills (Sun et al., 2020). According to the authors, this model was derived from synthesizing prior research on CPS and could guide CPS skills' assessment across multiple domains.

The considerable varieties in the theoretical models may reflect the diversity of assessment purposes of CPS skills. For example, the aim of the model proposed by ATC21S is to develop an assessment which supports peer interactions via online chat box (Hesse et al., 2015; Scoular, Care, & Hesse, 2017), while the model from PISA is intended to construct summative assessment in H-A mode (Rojas et al., 2021; von Davier et al., 2017). Another case was from Sun et al. (2020), who focused on authentic H–H interaction where three participants in a team assumed different roles, either assigned or naturally. Notably, the lack of consensus on the theoretical model of CPS skills would result in a challenge of accurately assessing CPS skills and comparing the research findings. As indicated by Sun et al. (2020), without a consensus on this issue, it would be impossible to evaluate them effectively and promote the generalizability of study results. Therefore, there is a great call for a consensus on the theoretical model of CPS skills. Sun et al. (2020) made an attempt in this stream of research by synthesizing prior research on CPS skills to construct their new model, i.e., the generalized competence model of CPS skills. This model was claimed to be applicable for diverse assessment purposes. However, its validity has not been comprehensively tested. As such, more validation research should be conducted on existing or newly-developed synthesized models for various assessment purposes.

Besides, we found that the theoretical models have been used in diversified assessment contexts regarding participants' age levels, collaboration features, and team compositions. It would thus be valuable to set up a searchable database where the theoretical models are systematically linked to their relevant assessment contexts. By doing so, it could be convenient for both researchers and practitioners to identify the optimal models for their assessments of CPS skills. However, most of these models were used across a narrow range of participants' age levels, collaboration features, and team compositions. That is, none model has been used across all assessment contexts, even the widespread ones from PISA and ATC21S. For example, the model from PISA has not been adopted in the assessment context supported by face-to-face collaboration in H–H mode; Co-measure has been only used for assessing CPS skills of elementary students (Herro et al., 2021). Therefore, an interesting research question arises: Is each theoretical model specific to a certain assessment context or suitable for all ones? A careful examination of this question would help determine the assessment contexts to which each theoretical model could apply.

4.3. Assessments of CPS skills and their characteristics

Our analysis revealed that a total of twenty-two computer-based tests were adopted for assessing CPS skills, and most of them fell into the first two categories in our constructed classification framework, i.e., tests with specific predefined messages in H-A mode and

online chat box in H–H mode. In contrast, tests with videoconferencing and face-to-face collaboration in H–H mode (quite similar to authentic assessment) were under-developed and –utilized. Only seven studies employed these two types of assessments to assess participants' CPS skills. This finding might be due to the difficulty of constructing an appropriate assessment environment for these two types of assessments. According to the characteristics of CPS skills, they should be assessed in an environment where participants cannot achieve the goal of CPS task until they work collaboratively to solve the problems in the assessments (Graesser et al., 2018). As summarized by Rojas et al. (2021), a good collaborative condition for assessing CPS skills should include positive interdependence between participants, individual accountability, and awareness of peers' work, etc. However, it is usually difficult to meet these requirements in the assessment contexts supported by videoconferencing or face-to-face collaboration in H–H mode. A compensatory solution is to assign different roles for each participant in the collaborative team and make them take on different responsibilities for completing the CPS task(s). For example, Sun et al. (2020) specifically assigned a particular role randomly for each student in triads at the beginning of a game play and switched their roles each time they reached a new game level; In the study of Stewart et al. (2021), students were randomly assigned with certain fixed roles throughout the process of executing the assigned CPS task. Given the increasing call for authentic assessment of CPS skills (Nouri et al., 2017), future research should make an attempt to develop more appropriate tests with videoconferencing and face-to-face collaboration in H–H mode and incorporate role assignment approach into the design of assessments.

Besides, there were certain differences among all assessment types in terms of participants' age levels and types of CPS task(s). Regarding participants' age levels, all tests were adopted for a limited range except *PISA 2015 CPS* and *ATC21S CPS*. These findings indicate that more studies should be conducted to test the applicability of these tests across various levels of participants' age. Concerning types of CPS task(s), most tests with specific predefined messages in H-A mode employed content-free task(s), while the remaining three ones mainly employed content-dependent task(s). These findings might be attributed to the differences of assessment purposes in these tests. As indicated by Griffin and Care (2015), content-free task focuses on students' hypothetico-deductive reasoning skills, whereas content-dependent task draws on particular skills and knowledge derived from certain curriculum-based contents. Noticeably, we also found that the content-dependent tasks in the assessment were mainly related with science curriculums. Only two studies introduced course contents in language and literacy and information technology in the assessment. Some studies have found that students could demonstrate their CPS skills in subject domains like business and economics and educational technology (e.g., Ouyang, Ling, & Jiao, 2021; Paeβens & Winther, 2021). Thus, future studies could attempt to embed some non-science contents into the assessments to expand CPS skill assessing to more subject domains.

Furthermore, our proposed four types of tests demonstrated great diversities in team compositions, types of assessment data, and methods of data recording and scoring. For tests with specific predefined messages in H-A mode, their commonly-used team composition was a team of one human being assessed and one to three virtual agents. Besides, the assessment data produced by them were numerous specific predefined messages, which were recorded by log-file and scored automatically by the computer. All these features equip this type of assessments with the advantages of convenient administration, efficient, standardized, and large-scale data collection, assessing a wider spectrum of CPS skills (Rosen, 2015), and a good control of external effects, such as group composition effects (Chen & Kuo, 2019) and personality effects (Herborn, Mustafic, & Greiff, 2018). However, they are often criticized for their limitation of employing predefined messages in a range of unnatural CPS activities, as well as a lack of discourse mechanisms that are core to CPS. In addition, this type of assessments might not be the best way to assess CPS skills as it does not consider the personality of the team members and their emotions, which have been proved to be important psychological factors for CPS skills (Graesser, Greiff, Stadler, & Shubeck, 2020). Therefore, future research is needed to promote the ecological validity of tests in this type. According to Pásztor-Kovács, Pásztor, and Molnár (2021), a good solution is to design a H–H pre-version that permits open-ended discussions prior to designing certain number of realistic and intelligent agents. Through an in-depth analysis of the data collected from the H–H version, it is possible for assessment developer to create well-established predefined messages.

For test with online chat box in H–H mode, they mainly employed the collaborative team composed of two humans. Besides, their generated assessment data were actions and chats, which were recorded by log files. These features make this type of assessments superior in providing more natural human collaboration situations where participants were allowed to type unexpected individual messages or take unnatural actions during their collaborations (e.g., Herborn et al., 2020; Stadler et al., 2020). As to the method of data scoring, the data generated by most tests were scored manually. One possible explanation is that it is difficult to automatically score the unexpected messages and unnatural actions by standardized algorithms. Although Adams et al. (2015) provided an approach for automatically scoring the assessment data for *ATC21S CPS*, only the data of actions, but not chats, could be addressed. Yuan et al.'s (2019) proposed automatic scoring approach was grounded on manual coding of a few students' chats. Considering the costly and time-consuming nature of human coding, we strongly suggest implementing large-scale tests of assessments across various samples. Through multiple iterations, it is possible for future researchers to constantly optimize their developed automatic scoring programs until that all chats and actions from their assessments could be scored automatically.

Remarkably, great diverseness was found in the team compositions for executing the CPS tasks in this type of assessments. This finding reflects researchers' debate on the optimal team size for collaborative activities. Currently, a team of two participants occupies the maximum proportion of team compositions. It has, however, faced a wave of criticism for the fact that participants in dyad are highly dependent on each other. If one participant in each dyad has bad performance, the other is unlikely to obtain a good score of CPS skills. As indicated by Pásztor-Kovács et al. (2021), participants are more likely to manifest their CPS skills in a team with multiple members with different abilities (Hao, Liu, von Davier, & Kyllonen, 2017; Rosen, 2017). Yet, the larger size of the CPS team would add difficulty to the development of assessments in H–H mode as positive interdependence between participants is highly emphasized in designing CPS tasks (Rojas et al., 2021). Therefore, assessment developers have to make a balance for the number of participants in the CPS team. According to some researchers, triads are recommended for CPS tasks (Zurita, Nussbaum, & Salinas, 2005). And, this team

size has been proved to promote negotiation and debate among team members (Nussbaum et al., 2009). However, it remains a possibility that in each team, three participants who do not excel in CPS tasks are grouped together and thus unable to achieve satisfactory performance. In addition, the adoption of teams composed of three participants will definitely increase difficulty for designing CPS tasks appropriate for providing sufficient opportunities for all three participants to demonstrate their CPS skills. Therefore, future research should explore how to develop an optimized team formation scheme accompanied by developing more appropriate CPS tasks for triads to improve the tests of CPS skills. In the last few years, there has been an interest to study group formation scheme that considers numerous student characteristics to optimize collaborative learning groups (e.g., Chen & Kuo, 2019; Sadeghi & Kardan, 2015). Given the considerable similarities between collaborative learning and CPS activities, it is of great value to draw on or improve existing group formation schemes to form optimal collaborative teams in assessments of CPS skills.

For the test with videoconferencing in H–H mode, it was conducted in the team consisting of three humans. Apart from actions and chats, facial expressions were also the assessment data it could provide. In addition, the researchers videotaped the assessment data, which were scored manually in most studies. However, our in-depth examination showed that only chats were taken into consideration in the manual scoring process (Stewart et al., 2021; Sun et al., 2020). Apparently, this practice would lose much important information about participants' CPS performance during data scoring. To address this gap, Stewart et al. (2021) made the first attempt to model participants' CPS skills from multimodal data of linguistic, task context, facial expressions, and acoustic–prosodic features by both standard and deep sequential learning classifiers. Considering the lack of multimodal information for data scoring, future studies are needed to model participants' CPS skills from as much non-redundant information as possible in the process of data scoring.

For tests with face-to-face collaboration in H–H mode, they adopted the team of at least three humans and could provide the unique data of physical interactions apart from the data types as mentioned earlier. In addition, video recording was adopted for recording the assessment data, and human coding was for data scoring. However, existing scoring rubrics focus more on the chats, while the additional information of physical interactions is largely overlooked. Researchers have demonstrated that students' physical interactions could be well used to predict their CPS skills (Cukurova, Luckin, Millán, & Mavrikis, 2018). Therefore, it is necessary to incorporate the information of physical interactions into the scoring rubrics to gain a comprehensive evaluation of participants' CPS skills.

4.4. Reliability and validity evidence of assessments of CPS skills

Our review found that a certain number of the reviewed studies did not provided the quality evidence of reliability and validity. As reliability and validity information are excellent proxies of assessment quality, a lack of their detailed results would lead to researchers' and educators' hesitations in adopting and promoting certain CBA of CPS skills. In addition, without sufficient reporting of both reliability and validity arguments, it is impossible for researchers to make appropriate revisions when improving or developing assessments of CPS skills. Therefore, more research is needed to establish comprehensive reliability and validity evidence for the assessments of CPS skills. Although most of the selected studies specifically investigated the psychometric properties of certain tests, they did not cover all evidence and analyses for triangulating the corresponding tests of CPS skills.

As mentioned in many studies (e.g., Messick, 1995; Siddiq, Hatlevik, Olsen, Throndsen, & Scherer, 2016), researchers have proposed some well-constructed theories or frameworks to provide practical standards for the constituents of validity evidence of assessments. One important problem is that these standards do not yet provide detailed and general information on how to conduct validation research (Shaw & Hughes, 2015). Therefore, relevant researchers may have to specifically match the process of validation to the utilized test, the examined sample, and the adopted theoretical model. However, this process of validation may be impeded for a lack of clear principles. Another negative consequence may be embodied in more time, money, and attention taken in validation studies.

5. Conclusions and suggestions

This literature review systematically summarizes the current state-of-the-art of CBA of CPS skills and points out future research directions to assess these skills appropriately and accurately.

The results revealed that CBA instruments of CPS skills were developed and adopted unequally in research contexts. They were more frequently employed in a sample of less than 500 junior students. A variety of theoretical models were adopted for assessing CPS skills, of which most studies adopted the separated model that separated the cognitive and social aspects of CPS skills. In addition, the majority of these theoretical models were applied to a limited range of participants' age levels, collaboration features, and team compositions. Four types of CBA instruments of CPS skills emerged in the selected studies: tests with specific predefined messages in H-A mode, and those with online chat box, videoconferencing, and face-to-face collaboration in H–H mode. They were more frequently adopted in a narrow range of participants' age levels. The majority of studies employed the first two types of assessments to assess CPS skills. Most tests of the first type employed content-free task(s), while the remaining three ones mainly employed content-dependent task(s). Furthermore, tests of the first type were more conducted in a team of one human being assessed and one to three virtual agents. The assessment data produced by them were numerous specific predefined messages, which were recorded by log-file and scored automatically by the computer. Tests of the second type mainly employed the collaborative team of two humans, and their generated assessment data were actions and chats, which were recorded by log files. And, the data generated by most tests were scored manually. Tests of the third type was implemented in the team of three humans. It could provide the assessment data of actions, chats and facial expressions, which were scored manually in most studies. Tests of the last type adopted the team of at least three humans and could provide the unique data of physical interactions. Researchers videotaped the assessment data and manually coded them. Lastly, a

considerable number of studies failed to provide the reliability and validity evidence of their assessments.

Despite great work from the reviewed studies, more research is still needed to spur the development of assessments of CPS skills. In particular, we recommend that researchers and assessment developers on CPS skills should take into account the following tips when designing and reporting CBA instruments of CPS skills: (a) creating more tests of CPS skills for participants at the age levels of elementary and senior student and adult; (b) examining whether each theoretical model is specific to a certain assessment context or suitable for all ones; (c) validating existing synthesized theoretical models and developing new ones to assess CPS skills across various assessment purposes; (d) developing more tests with videoconferencing and face-to-face collaboration in H-A mode; (e) expanding tests of CPS skills to non-science subject domains; (f) developing tests with specific predefined messages in H-A mode by designing a H–H pre-version prior to creating certain number of realistic and intelligent agents; (g) implementing large-scale tests of assessments with online chat box in H–H mode across various samples to optimize automatic scoring programs; (h) optimizing team formation accompanied by designing appropriate CPS tasks for triads in tests with online chat box in H–H mode; (i) modeling participants' CPS skills from multimodal information in tests with videoconferencing in H–H mode; (j) incorporating the information of physical interactions into the scoring rubrics for tests with face-to-face collaboration in H–H mode; (g) providing reliability and validity evidence to confidently qualify the test.

Finally, given the complex and diverse nature of CPS skills, researchers and practitioners across different disciplines and fields should conduct in-depth collaboration to assess CPS skill comprehensively and systematically. Only in this way, can CPS skills be captured appropriately and further be well cultivated.

Author statement

Huanyou Chai: Conceptualization, Methodology, Writing- Original draft preparation, Writing- Reviewing and Editing, Formal analysis. Tianhui Hu: Methodology, Writing- Original draft preparation, Writing- Reviewing and Editing, Formal analysis. Li Wu: Writing- Reviewing and Editing, Formal analysis.

Data availability

Data will be made available on request.

Acknowledgement

This work was supported by National Social Science Foundation of China [Project No. CCA230340].

Appendix A

The Reviewed Studies on the Assessment of Collaborative Problem Solving Skills.

Author (year)	Theoretical model	Age level	Sample size	Test or task name	Assessment environment	Assessment mode	Task type	Collaboration medium	Group composition	Country/ Region
Ahonen and Harding (2018)	ATC21SM	Age = 11-15	480	*ATC21S CPS*	Online	H–H	Both	OCB	2Hs	ATC21S; *Finland
Amon, Vrzakova and D'Mello (2019)	GNEM	University; age = 19.40	64	*Minecraft Hour of Code*	Online	H–H	DEP	VIF	3Hs	*USA
Andrews-Todd and Forsyth (2020)	ONTM	age = 16-60	129	*Three-Resistor Activity*	Online	H–H	DEP	OCB	3Hs	*USA
Azura et al.(2021) a	AZUM	Age = 16	30	A physics CPS task	Online	H–H	DEP	OCB	3Hs	*Indonesia
Camacho-Morles et al.(2019)	ATC21SM	Junior; age = 13.48	22	*ATC21S CPS*	Online	H–H	Both	OCB	2Hs	ATC21S; *Australia
De Boeck & Scalise (2019)	PISAM	Age = 25	994	*PISA 2015 CPS*	Online	H-A	FRE	SPMs	1H/(1–3)As	OECD; USA
Dowell et al. (2020)	LIUM	Age = 18-68	967	*ATC21S CPS*	Online	H–H	DEP	OCB	2Hs	*USA
Ham and Hwang (2021)	PISAM	Junior; age = 15	11109	*PISA 2015 CPS*	Online	H-A	FRE	SPMs	1H/(1–3)As	OECD; USA, Korean
Hao et al. (2016)	LIUM	Adults	878	A collaborative simulation task	Online	H–H	DEP	OCB	2Hs	*USA
Hao et al. (2015)	LIUM	Adults	556	A collaborative simulation task	Online	H–H	DEP	OCB	2Hs	*USA

(*continued on next page*)

(*continued*)

Author (year)	Theoretical model	Age level	Sample size	Test or task name	Assessment environment	Assessment mode	Task type	Collaboration medium	Group composition	Country/ Region
Harding et al. (2017)	ATC21SM	Age = 11-17	3004	*ATC21S CPS*	Online	H–H	DEP	OCB	2Hs	ATC21S; *Australia
Herborn et al. (2020)	PISAM	Grade = 9–10; age = 15.60	748	*PISA 2015 CPS*	Online	H-A	FRE	SPMs	1H/(1–3)As	OECD; Germany
Herro et al. (2021)	Co-measure	Elementary; age = 6-10	52	Makerspace activities	FtF	H–H	DEP	FAC	Dynamic	*USA
Krkovic et al. (2016)	KRKM	Grade 7	483	A self-development assessment (a)	Online	H-A	FRE	SPMs	1H/1A	Luxembourg; Germany
Kuo et al. (2020)	PISAM	Grade = 9-10	53855	*PISA 2015 CPS*	Online	H-A	FRE	SPMs	1H/(1–3)As	OECD; Taiwan (China)
Lin et al. (2020)	PISAM	Junior	241	Assessment System for CPS skills	Online	H-A	FRE	SPMs	1H/(1–2)As	*Taiwan (China)
Lin et al. (2015)	PISAM	Junior; age = 13-15	222	Assessment System for CPS skills	Online	H-A	FRE	SPMs	1H/(1–2)As	*Taiwan (China)
Lin et al. (2021)	ONTM	University	525	*Collaborative Science Assessment Prototype*	Online	H–H	DEP	OCB	(3–5)Hs	*USA
Nouri et al. (2017)	PISAM	Junior; age = 13-14	24	A balancing scale problem task	Online	H–H	DEP	OCB	2Hs	*Sweden
Polyak et al. (2017)	HOLF	Junior	500	*Circuit Runner*	Online	H-A	FRE	SPMs	1H/1A	*USA
Pöysä-Tarhonen et al.(2	ATC21SM	Master-level teacher	20	*ATC21S CPS*	Online	H–H	Both	OCB	2Hs	ATC21S; *Finland

Author (year)	Theoretical model	Age level	Sample size	Test or task name	Assessment environment	Assessment mode	Task type	Collaboration medium	Group composition	Country/ Region
Pöysä-Tarhonen et al.(2021)	ATC21SM	Age = 27	15	Virtual manipulatives	FtF	H–H	DEP	FAC	(3–4)Hs	*Finland
Rojas et al. (2021)	PISAM	Elementary; age = 10-13	719	A constellation or molecule building task	Online	H-A	FRE	SPMs	1H/2As	*Chile
Rosen (2017)	PISAM	Junior; age = 14	220	*Zoo Quest*	Online	H-A	DEP	SPMs	1H/1A	*USA
Rosen (2015)	PISAM	Junior; age = 14	179	*Zoo Quest*	Online	H-A	DEP	SPMs	1H/1A	*USA
Rosen et al. (2020)	HOLF	Grade = 6-9	196	*Science Fair*	Online	H-A	DEP	SPMs	No info	*USA
Scoular and Care (2020)	ATC21SM	Age = 12-15	3010	*ATC21S CPS*; A self-development assessment (b)	Online	H–H	Both	OCB	2Hs	ATC21S; *Australia
Song and Lan (2018)	PISAM	Elementary; grade = 6	53	A biology CPS task	FtF	H–H	DEP	FAC	6Hs	*HK (China)
Song et al. (2020)	PISAM	University	56	*Concept Map*	Online	H–H	DEP	OCB	(3–4)Hs	*Korean
Stadler et al. (2020)	PISAM	Grade = 9-10	483	*PISA 2015 CPS*	Online	H-A	FRE	SPMs	1H/(1–3)As	OECD; Germany
Stewart et al. (2021)	GNEM	University; age = 19.4	111	*Minecraft Hour of Code*	Online	H–H	DEP	VIF	3Hs	*USA
Stoeffler et al., 2020	HOLF	Age = 33.76	379	*Circuit Runner*	Online	H-A	FRE	SPMs	1H/1A	*USA
Sun et al. (2020)	GNEM	Age = 14-15	33	*Physics Playground*	FtF	H–H	DEP	FAC	3Hs	*USA
		Age = 19.4	111	*Minecraft Hour of Code*	Online	H–H	DEP	VIF	3Hs	*USA
Tang et al. (2021)	PISAM	Junior; age = 15	9841	*PISA 2015 CPS*	Online	H-A	FRE	SPMs	1H/(1–3)As	OECD; China
Tekin and Aktan (2021)	PISAM	Junior	2990	*PISA 2015 CPS*	Online	H-A	FRE	SPMs	1H/3As	OECD; Singapore, Turkey, and Norway
Tsang et al. (2020)	ATC21SM	Elementary; grade = 3	34	Problems related to cyberbullying	Online	H–H	FRE	OCB	(3–5)Hs	*HK (China)

(*continued on next page*)

(*continued*)

Author (year)	Theoretical model	Age level	Sample size	Test or task name	Assessment environment	Assessment mode	Task type	Collaboration medium	Group composition	Country/ Region
Tsang et al. (2019)	ATC21SM	Age = 11-15	44	learning and game design *ATC21S CPS*; Problems related to cyberbullying learning and game design	Online	H–H	Both	OCB	2Hs, (3–5) Hs	*HK (China)
Wang and Hu (2021)	PISAM	Junior; age = 15	46268	*PISA 2015 CPS*	Online	H-A	FRE	SPMs	1H/(1–3)As	OECD; China
Yavuz and Atar (2020)	PISAM	Age = 15	435	*PISA 2015 CPS*	Online	H-A	FRE	SPMs	1H/(1–3)As	OECD; Turkey
Yuan et al. (2019)	ATC21SM	Junior; age = 15	434	A self-development assessment (c)	Online	H–H	FRE	OCB	2Hs	*China

Note.
1.In the category *theoretical model*, the following abbreviations have been used: GNEM = The generalized competence model, HOLF = The Holistic framework, PISAM = The model proposed by PISA, ATC21SM = The model proposed by ATC21S, ONTM = The model from a CPS ontology, AZUM = The model by Azura et al.(2021)a, KRKM = The model by Krkovic et al. (2016), LIUM = The model by Liu et al. (2016).
2. In the category *Test name or task*, letters in the parentheses after test name or task were used to distinguish between different assessment ones.
3. In the category *Assessment environment*, Online = Online environment, FtF = Face-to-Face environment.
4. In the category *Task type*, DEP = Content-dependent task(s); FRE = content-free task(s); 'Both' represents that DEP and FRE were all adopted.
5. In the category *Collaboration medium*, SPMs = Specific predefined messages, OCB = Online chat box, VIF = videoconferencing, FAC = Face-to-face collaboration.
6. n the category *Team composition*, H= Human, A = Agent.
7. In the category *Country/Region*, the first name refers to the country or international organization in which the assessment was developed. The asterisk (*) indicates that the assessment was also administered in this country.

Appendix B

An overview of theoretical models in the '*separated*' model type for assessing CPS skills in the reviewed studies.

Theoretical model	Dimensions			
The model by PISA	Social Cognitive	(1) Establishing and maintaining shared understanding	(2) Taking appropriate action to solve the problem	(3) Establishing and maintaining team organization
	(A) Exploring and understanding	(A1) Discovering perspectives and abilities of team members	(A2) Discovering the type of collaborative interaction to solve the problem, along with goals	(A3) Understanding roles to solve problem
	(B) Representing and formulating	(B1) Building a shared representation and negotiating the meaning of the problem (common ground)	(B2) Identifying and describing tasks to be completed	(B3) Describe roles and team organization (communication protocol/rules of engagement)
	(C) Planning and executing	(C1) Communicating with team members about the actions to be/ being performed	(C2) Enacting plans	(C3) Following rules of engagement, (e.g., prompting other team members to perform their tasks.)
	(D) Monitoring and reflecting	(D1) Monitoring and repairing the shared understanding	(D2) Monitoring results of actions and evaluating success in solving the problem	(D3) Monitoring, providing feedback and adapting the team organization and roles
The model by ATC21S	**Social**: Participation, perspective taking, and social regulation skills; **Cognitive**: Learning and knowledge building and task regulation skills			
Liu et al.'s (2016) model	**Social**: Sharing ideas, negotiating ideas, regulating problem-solving activities, and maintaining communication; **Cognitive**: Conceptual understanding and inquiry skills			
The Holistic framework	**Social**: Inclusiveness, clarity, communication, and commitment; **Cognitive**: Problem orientation, goal orientation, strategy, execution, and monitoring and evaluating			
The model from an ontology	**Social:** Maintaining communication, sharing information, establishing shared understanding, and negotiating; **Cognitive**: exploring and understanding, representing and formulating, planning, executing, and monitoring			
Co-measure	**Social**: Peer interactions and positive communication; **Cognitive**: Inquiry rich/multiple paths and transdisciplinary approaches			
Azura et al.'s model (2021)	**Social**: Participation, perspective taking, and social regulation skills; **Cognitive**: Knowledge building and task regulation skills			
Krkovic et al.'s model (2016)	**Social**: Questioning, asserting, and requesting; **Cognitive**: Knowledge acquisition and knowledge application			

References

Adams, R., Vista, A., Scoular, C., Awwal, N., Griffin, P., & Care, E. (2015). Automatic coding procedures for collaborative problem solving. In P. Griffin, & E. Care (Eds.), *Assessment and teaching of 21st century skills: Methods and approach* (pp. 115–132). Springer.

* Ahonen, A. K., & Harding, S. M. (2018). Assessing online collaborative problem solving among school children in Finland: A case study using ATC21S TM in a national context. *International Journal of Learning, Teaching and Educational Research, 17*(2), 138–158.

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

* Andrews-Todd, J., & Forsyth, C. M. (2020). Exploring social and cognitive dimensions of collaborative problem solving in an open online simulation-based task. *Computers in Human Behavior, 104*, Article 105759. https://doi.org/10.1016/j.chb.2018.10.025.

Andrews-Todd, J., & Kerr, D. (2019). Application of ontologies for assessing collaborative problem solving skills. *International Journal of Testing, 19*(2), 172–187. https://doi.org/10.1080/15305058.2019.1573823

Aqlan, F., & Zhao, R. (2022). Assessment of collaborative problem solving in engineering students through hands-on simulations. *IEEE Transactions on Education, 65*(1), 9–17. https://doi.org/10.1109/TE.2021.3079523

Autor, D. H., Levy, F., & Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *Quarterly Journal of Economics, 118*(4), 1279–1333. https://doi.org/10.1162/003355303322552801

* Azura, A., Suwarma, I. R., & Efendi, R. (2021a). Measuring collaborative problem-solving skills (CPSS) of vocational high school students using web-based assessment. *Journal of Physics: Conference Series, 1806*(1), 12–25. https://doi.org/10.1088/1742-6596/1806/1/012025.

Baligar, P., Joshi, G., Shettar, A., Guerra, A., Kolmos, A., Chen, J., et al. (2020). Assessment of collaborative-problem solving competency in engineering education: A systematic literature review, 2020-01-01. In *Paper presented as the 8th international research symposium on problem-based learning* (pp. 592–604).

Bland, L. M., & Gareis, C. R. (2018). Performance assessments: A review of definitions, quality characteristics, and outcomes associated with their use in K-12 schools. *Teacher Educators' Journal, 11*, 52–69.

Burrus, J., Jackson, T., Xi, N., & Steinberg, J. (2013). *Identifying the most important 21st century workforce competencies: An analysis of the occupational Information network (O* NET) (No. ETS RR-13-21)*. Princeton, NJ: Educational Testing Service.

* Camacho-Morles, J., Slemp, G. R., Oades, L. G., Morrish, L., & Scoular, C. (2019). The role of achievement emotions in the collaborative problem-solving performance of adolescents. *Learning and Individual Differences, 70*, 169–181. https://doi.org/10.1016/j.lindif.2019.02.005.

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.

Care, E., Griffin, P., & Wilson, M. (2018). *Assessment and teaching of 21st century skills: Research and applications*. Cham: Springer.

Care, E., & Griffion, P. (2014). An approach to assessment of collaborative problem solving. *Research and Practice in Technology Enhanced Learning, 9*(3), 367–388.

Chadegani, A. A., Salehi, H., Yunus, M. M., Farhadi, H., Fooladi, M., Farhadi, M., et al. (2013). A comparison between two main academic literature collections: Web of Science and Scopus databases. *Asian Social Science, 9*(5), 18–26. https://doi.org/10.5539/ass.v9n5p18

Chen, C., & Kuo, C. (2019). An optimized group formation scheme to promote collaborative problem-based learning. *Computers & Education, 133*, 94–115.

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

von Davier, A. A., Hao, J., Liu, L., & Kyllonen, P. (2017). Interdisciplinary research agenda in support of assessment of collaborative problem solving: Lessons learned from developing a collaborative science assessment prototype. *Computers in Human Behavior, 76*, 631–640. https://doi.org/10.1016/j.chb.2017.04.059

De Boeck, P., & Scalise, K. (2019). Collaborative problem solving: Processing actions, time, and performance. *Frontiers in Psychology, 10*, 1280.

Dillenbourg, P., & Traum, D. (2006). Sharing solutions: Persistence and grounding in multimodal collaborative problem solving. *The Journal of the Learning Sciences, 15*(1), 121–151. https://doi.org/10.1207/s15327809jls1501_9

Dowell, N. M. M., Lin, Y., Godfrey, A., & Brooks, C. (2020). Exploring the relationship between emergent sociocognitive roles, collaborative problem-solving skills, and outcomes: a group communication analysis. *Journal of Learning Analytics, 7*(1), 38–58.

Dumontheil, I. (2014). Development of abstract thinking during childhood and adolescence: The role of rostrolateral prefrontal cortex. *Developmental cognitive neuroscience, 10*, 57–76. https://doi.org/10.1016/j.dcn.2014.07.009

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2015). *How to design and evaluate research in education* (9thed.). New York, NY: McGraw-Hill Education.

Fuad, A. Z., Alfin, J., Fauzan, Astutik, S., & Prahani, B. K. (2019). Group science learning model to improve collaborative problem solving skills and self-confidence of primary schools teacher candidates. *International Journal of Instruction, 12*(3), 119–132. https://doi.org/10.29333/iji.2019.1238a

Gonyea, R. M. (2005). Self-reported data in institutional research: Review and recommendations. In P. D. Umbach (Ed.), *New directions for institutional research* (Vol. 127, pp. 73–89). San Francisco: Jossey-Bass.

González-Betancor, S. M., Bolívar-Cruz, A., & Verano-Tacoronte, D. (2019). Self-assessment accuracy in higher education: The influence of gender and performance of university students. *Active Learning in Higher Education, 20*(2), 101–114. https://doi.org/10.1177/1469787417735604

Graesser, A. C., Fiore, S. M., Greiff, S., Andrews-Todd, J., Foltz, P. W., & Hesse, F. W. (2018). Advancing the science of collaborative problem solving. *Psychological Science in the Public Interest, 19*(2), 59–92. https://doi.org/10.1177/1529100618808244

Graesser, A. C., Greiff, S., Stadler, M., & Shubeck, K. T. (2020). Collaboration in the 21st century: The theory, assessment, and teaching of collaborative problem solving. *Computers in Human Behavior, 104*, Article 106134. https://doi.org/10.1016/j.chb.2019.09.010

Griffin, P., & Care, E. (2015). *Assessment and teaching of 21st century skills: Methods and approach*. Dordrecht: Springer.

Griffin, P., Care, E., & Harding, S. M. (2015). Task characteristics and calibration. In *Assessment and teaching of 21st century skills* (pp. 133–178). Dordrecht: Springer.

Griffin, P., McGaw, B., & Care, E. (2012). *Assessment and teaching of 21s century skills*. Dordrecht: Springer.

Gu, X., Chen, S., Zhu, W., & Lin, L. (2015). An intervention framework designed to develop the collaborative problem-solving skills of primary school students. *Educational Technology Research & Development, 63*(1), 143–159. https://doi.org/10.1007/s11423-014-9365-2

* Ham, Y., & Hwang, J. (2021). Mathematical literacy and collaborative problem-solving: Comparison between Korean and U.S. Students in PISA2015. *The Journal of Educational Research in Mathematics, 31*(3), 299–320. https://doi.org/10.29275/jerm.2021.31.3.299.

Hao, J., Liu, L., von Davier, A. A., & Kyllonen, P. C. (2017). Initial steps towards a standardized assessment for CPS: Practical challenges and strategies. In A. A. von Davier, M. Zhu, & P. C. Kyllonen (Eds.), *Innovative assessment of collaboration*. New York: Springer.

* Hao, J., Liu, L., von Davier, A. A., Kyllonen, P., & Kitchen, C. (2016). Collaborative problem solving skills versus collaboration outcomes: Findings from statistical analysis and data mining. In *Paper presented at the Proceedings of the 9th international conferences on educational data mining, Raleigh*. USA: North Carolina.

* Hao, J., Liu, L., von Davier, A., Kyllonen, P., Lindwall, O., Hakkinen, P., et al. (2015). Assessing collaborative problem solving with simulation based tasks, 2015-01-01. In *Paper presented at the 11th international conference on computer supported collaborative learning: Exploring the material conditions of learning*. CSCL 2015.

* Harding, S. E., Griffin, P. E., Awwal, N., Alom, B. M., & Scoular, C. (2017). Measuring collaborative problem solving using mathematics-based tasks. *AERA open, 3*(3). https://doi.org/10.1177/2332858417728046.

Herborn, K., Mustafic, M., & Greiff, S. (2018). *Computer-based collaborative problem solving in PISA 2015 and the role of the big five*. Manuscript submitted for publication.

* Herborn, K., Stadler, M., Mustafić, M., & Greiff, S. (2020). The assessment of collaborative problem solving in PISA 2015: Can computer agents replace humans?. *Computers in Human Behavior, 104*, Article 105624. https://doi.org/10.1016/j.chb.2018.07.035.

* Herro, D., Quigley, C., & Abimbade, O. (2021). Assessing elementary students' collaborative problem-solving in makerspace activities. *Information and Learning Sciences, 122*, 774–794. https://doi.org/10.1108/ils-08-2020-0176.

Herro, D., Quigley, C., Andrews, J., & Delacruz, G. (2017). Co-measure: Developing an assessment for student collaboration in STEAM activities. *International Journal of STEM Education, 4*(1), 26. https://doi.org/10.1186/s40594-017-0094-z

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 and teaching of 21st century skills: Methods and approach (37–56)*. Dordrecht: Springer. (Reprinted).

Hodes, L. N., & Thomas, K. G. F. (2020). Inaccuracy of self-reports and influence of psychological and contextual factors. *Computers in Human Behavior*. , Article 106616. https://doi.org/10.1016/j.chb.2020.106616

Holstein, J. A., & Gubrium, J. F. (2001). *Handbook of interview research: Context and method*. Thousand oaks, CA: Sage.

* Krkovic, K., Wüstenberg, S., Greiff, S., Roberts, R. D., Petway, K., & Allen, V. (2016). Assessing collaborative behavior in students: An experiment-based assessment approach. *European Journal of Psychological Assessment: Official Organ of the European Association of Psychological Assessment, 32*(1), 52–60. https://doi.org/10.1027/1015-5759/a000329.

* Kuo, B. C., Liao, C. H., Pai, K. C., Shih, S. C., Li, C. H., & Mok, M. M. C. (2020). Computer-based collaborative problem-solving assessment in Taiwan. *Educational Psychology, 40*(9), 1164–1185. https://doi.org/10.1080/01443410.2018.1549317.

* Lin, Y., Dowell, N., & Godfrey, A. (2021). Skills matter: Modeling the relationship between decision making processes and collaborative problem-solving skills during Hidden Profile Tasks, 2021-01-01. In *Paper presented at the 11th international Conference on learning Analytics and knowledge* (pp. 428–437). New York: LAK 2021.

* Lin, K., Yu, K., Hsiao, H., Chang, Y., & Chien, Y. (2020). Effects of web-based versus classroom-based STEM learning environments on the development of collaborative problem-solving skills in junior high school students. *International Journal of Technology and Design Education, 30*(1), 21–34. https://doi.org/10.1007/s10798-018-9488-6.

* Lin, K., Yu, K., Hsiao, H., Chu, Y., Chang, Y., … Chien, Y. (2015). Design of an assessment system for collaborative problem solving in STEM education. *Journal of Computers in Education, 2*(3), 301–322. https://doi.org/10.1007/s40692-015-0038-x.

Liu, L., Von Davier, A., Hao, J., Kyllonen, P., & Zapata-Rivera, D. (2016). A tough nut to crack: Measuring collaborative problem solving. In R. Yigal, S. Ferrara, & M. Mosharraf (Eds.), *Handbook of research on technology tools for real-world skill development* (pp. 344–359). Hershey, PA: IGI Global.

Mayer, R. E., & Wittrock, M. C. (2006). Problem solving. In P. A. Alexander, & P. H. Winne (Eds.), *Handbook of educational psychology* (2nd ed., pp. 287–304).

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performance as scientific inquiry into score meaning. *American Psychologist, 50*, 741–749. https://doi.org/10.1037/0003-066x.50.9.741

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group.. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine, 6*(7), Article 1000097. https://doi.org/10.1371/journal.pmed.1000097

Molnár, G., Greiff, S., & Csapó, B. (2013). Inductive reasoning, domain specific and complex problem solving: Relations and development. *Thinking Skills and Creativity, 9*, 35–45. https://doi.org/10.1016/j.tsc.2013.03.002

* Nouri, J., Åkerfeldt, A., Fors, U., & Selander, S. (2017). Assessing collaborative problem solving skills in technology-enhanced learning environments – the PISA framework and modes of communication. *International Journal of Emerging Technologies in Learning (iJET), 12*(4), 163–174. https://doi.org/10.3991/ijet.v12i04.6737.

Nussbaum, M., Alvarez, C., McFarlane, A., Gomez, F., Claro, S., & Radovic, D. (2009). Technology as small group face-to-face collaborative scaffolding. *Computers & Education, 52*(1), 147–153. https://doi.org/10.1016/j.compedu.2008.07.005

O'Leary, M., Scully, D., Karakolidis, A., & Pitsia, V. (2018). The state-of-the-art in digital technology-based assessment. *European Journal of Education, 53*(2), 160–175. https://doi.org/10.1111/ejed.12271

OECD, O. F. E. C. (2013). *PISA 2015 draft collaborative problem solving framework*. Paris: OECD. Retrieved from http://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Collaborative%20Problem%20Solving%20Framework%20.pdf.

OECD, O. F. E. C. (2017). *PISA 2015 results (volume V): Collaborative problem solving (131–188*. Paris: OECD Publishing.

Oliveri, M. E., Lawless, R., & Molloy, H. (2017). *A literature review on collaborative problem solving for college and workforce readiness. ETS GRE® board research report. ETS GRE®-17-03. (ETS research report RR-17-06)*. Retrieved from https://online.library.wiley.com/doi/pdf/10.1002/ets2.12133.

O'Neil, H. F., Chuang, S.-h. S., & Baker, E. L. (2010). Computer-based feedback for computer-based collaborative problem solving. In *Computer-based diagnostics and systematic analysis of knowledge* (pp. 261–279). Springer.

Ostrander, A., Bonner, D., Walton, J., Slavina, A., Ouverson, K., Kohl, A., et al. (2020). Evaluation of an intelligent team tutoring system for a collaborative two-person problem: Surveillance. *Computers in Human Behavior, 104*, Article 105873. https://doi.org/10.1016/j.chb.2019.01.006

Ouyang, F., Ling, T., & Jiao, P. (2021). Development of group cognition in online collaborative problem-solving processes. *Journal of Educational Computing Research, 60*(3), 599–630. https://doi.org/10.1177/07356331211047784

Paeßens, J., & Winther, E. (2021). Game design in financial literacy: Exploring design patterns for a collaborative and inclusive serious game from different perspectives. In *Game-based learning across the disciplines* (pp. 43–59). Cham: Springer.

Pásztor-Kovács, A., Pásztor, A., & Molnár, G. (2021). Measuring collaborative problem solving: Research agenda and assessment instrument. *Interactive Learning Environments*, 1–21. https://doi.org/10.1080/10494820.2021.1999273

Perry, K., Meissel, K., & Hill, M. F. (2022). Rebooting assessment. Exploring the challenges and benefits of shifting from penand-paper to computer in summative assessment. *Educational Research Review, 36*, Article 100451. https://doi.org/10.1016/j.edurev.2022.100451

* Polyak, S. T., von Davier, A. A., & Peterschmidt, K. (2017). Computational psychometrics for the measurement of collaborative problem solving skills. *Frontiers in Psychology, 8*, 2029. https://doi.org/10.3389/fpsyg.2017.02029.

* Pöysä-Tarhonen, J., Häkkinen, P., Tarhonen, P., Näykki, P., & Järvelä, S. (2021). "Anything taking shape?" Capturing various layers of small group collaborative problem solving in an experiential geometry course in initial teacher education. *Instructional Science, 50*, 1–34. https://doi.org/10.1007/s11251-021-09562-5.

Putri, I. E., & Sinaga, P. (2021). Collaborative problem-solving: How to implement and measure it in science teaching and learning. *Journal of Physics: Conference Series, 1806*(1), 12–18. https://doi.org/10.1088/1742-6596/1806/1/012018

* Rojas, M., Nussbaum, M., Chiuminatto, P., Guerrero, O., Greiff, S., Krieger, F., et al. (2021). Assessing collaborative problem-solving skills among elementary school students. *Computers & Education, 175*, Article 104313. https://doi.org/10.1016/j.compedu.2021.104313.

Roschelle, J., & Teasley, S. (1995). The construction of shared knowledge in collaborative problem solving. In C. O'Malley (Ed.), *Computer-supported collaborative learning* (pp. 69–97). New York: Springer-Verlag.

* Rosen, Y. (2015). Computer-based assessment of collaborative problem solving: Exploring the feasibility of human-to-agent approach. *International Journal of Artificial Intelligence in Education, 25*(3), 380–406. https://doi.org/10.1007/s40593-015-0042-3.

* Rosen, Y. (2017). Assessing students in human-to-agent settings to inform collaborative problem-solving learning. *Journal of Educational Measurement, 54*(1), 36–53. https://doi.org/10.1111/jedm.12131.

Rosen, Y., Wolf, I., & Stoeffler, K. (2020). Fostering collaborative problem solving skills in science: The Animalia project. *Computers in Human Behavior, 104*, Article 105922. https://doi.org/10.1016/j.chb.2019.02.018

Rummel, N., & Spada, H. (2005). Learning to collaborate: An instructional approach to promoting collaborative problem solving in computer-mediated settings. *The Journal of the Learning Sciences, 14*(2), 201–241. https://doi.org/10.1207/s15327809jls1402_2

Sadeghi, H., & Kardan, A. A. (2015). A novel justice-based linear model for optimal learner group formation in computer-supported collaborative learning environments. *Computers in Human Behavior, 48*, 436–447. https://doi.org/10.1016/j.chb.2015.01.020

Schacter, J., Herl, H. E., Chung, G. K. W. K., Dennis, R. A., & O'Neil, H. F., Jr. (1999). Computer-based performance assessments: A solution to the narrow measurement and reporting of problem-solving. *Computers in Human Behavior, 15*, 403–418.

Scherer, R., Greiff, S., & Kirschner, P. A. (2017). Editorial to the special issue: Current innovations in computer-based assessments. *Computers in Human Behavior, 76*, 604–606. https://doi.org/10.1016/j.chb.2017.08.020

* Scoular, C., & Care, E. (2020). Monitoring patterns of social and cognitive student behaviors in online collaborative problem solving assessments. *Computers in Human Behavior, 104*, Article 105874. https://doi.org/10.1016/j.chb.2019.01.007.

Scoular, C., Care, E., & Hesse, F. W. (2017). Designs for operationalizing collaborative problem solving for automated assessment. *Journal of Educational Measurement, 54*(1), 12–35. https://doi.org/10.1111/jedm.12130

Shaffer, D. R., & Kipp, K. (2009). *Development psychology: Childhood and adolescence*. Brooks/Cole.
Shaw, S., & Hughes, S. (2015). Issues around how best to provide evidence for assessment validity: The challenge of validation. In *Pre-conference workshop at the association for educational assessment - Europe, glasgow*. Retrieved http://www.aea-europe.net/images/1_Conferences/Glasgow_2015/workshop_template_2015_SDS_SH.pdf. (Accessed 10 December 2015).
Siddiq, F., Hatlevik, O. E., Olsen, R. V., Throndsen, I., & Scherer, R. (2016). Taking a future perspective by learning from the past – a systematic review of assessment instruments that aim to measure primary and secondary school students' ICT literacy. *Educational Research Review, 19*, 58–84. https://doi.org/10.1016/j.edurev.2016.05.002
Slavin, R. E. (2017). Instruction based on cooperative learning. In *Handbook of research on learning and instruction* (pp. 388–404). New York, NY: Routledge Press.
* Song, Y., & Lan, Y. (2018). Improving primary students' collaborative problem solving competency in project-based science learning with productive failure instructional design in a seamless learning environment. *Educational Technology Research & Development, 66*(4), 979–1008. https://doi.org/10.1007/s11423-018-9600-3.
* Song, M. H., Park, J. A., & Park, J. (2020). Measuring collaborative problem solving capability in creative problem solving situation, 2020-01-01. In *Paper presented at the 21st ACM international conference on supporting group work*. GROUP, 2020.
* Stadler, M., Herborn, K., Mustafić, M., & Greiff, S. (2020). The assessment of collaborative problem solving in PISA 2015: An investigation of the validity of the PISA 2015 CPS tasks. *Computers & Education, 157*, Article 103964. https://doi.org/10.1016/j.compedu.2020.103964.
Stewart, A., & D'Mello, S. K. (2018). Connecting the dots towards collaborative AIED: Linking group makeup to process to learning. In *International conference on artificial intelligence in education* (pp. 545–556). Cham: Springer.
* Stewart, A. E. B., Keirn, Z., & D'Mello, S. K. (2021). Multimodal modeling of collaborative problem-solving facets in triads. *User Modeling and User-Adapted Interaction, 31*(4), 713–751. https://doi.org/10.1007/s11257-021-09290-y.
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.
* Sun, C., Shute, V. J., Stewart, A., Yonehiro, J., Duran, N., & D'Mello, S. (2020). Towards a generalized competency model of collaborative problem solving. *Computers & Education, 143*, Article 103672. https://doi.org/10.1016/j.compedu.2019.103672.
* Tang, P., Liu, H., & Wen, H. (2021). Factors predicting collaborative problem solving: Based on the data from PISA 2015. *Frontiers in Education, 6*, 130. https://doi.org/10.3389/feduc.2021.619450.
Taylor, K., & Baek, Y. (2018). Collaborative robotics, more than just working in groups. *Journal of Educational Computing Research, 56*(7), 979–1004. https://doi.org/10.1177/0735633117731382
* Tekin, Y. T., & Aktan, D. C. (2021). Investigation of measurement invariance of PISA 2015 collaborative problem solving skills: Turkey, Norway and Singapore. *International Journal of Assessment Tools in Education, 8*(1), 90–105. https://doi.org/10.21449/ijate.690576.
Timmis, S., Broadfoot, P., Sutherland, R., & Oldfield, A. (2016). Rethinking assessment in a digital age: Opportunities, challenges and risks. *British Educational Research Journal, 42*(3), 454–476. https://doi.org/10.1002/berj.3215
Tsang, H. W. C., Liu, Y., Ying Law, N. W., Gresalfi, M., & Horn, I. S. (2020, 2020/1/1). An in-depth study of assessment of collaborative problem solving (CPS) skills of students in both technological and authentic learning settings. In *Paper presented at the 14th International Conference of the Learning Sciences: The Interdisciplinarity of the Learning Sciences, ICLS 2020*.
Tsang, H. W. C., Won Park, S., Chen, L. L., Law, N. W. Y., Lund, K., Niccolai, G. P., et al. (2019, 2019/1/1). Assessing collaborative problem solving: What and how?. In *Paper presented at the 13th International Conference on Computer Supported Collaborative Learning - A Wide Lens: Combining Embodied, Enactive, Extended, and Embedded Learning in Collaborative Settings, CSCL 2019*.
Wang, M., & Hu, J. (2021, 2021/1/1). The influence of ICT-based social media on Asian students' collaborative problem-solving performance. In *Paper presented at the 16th IEEE International Conference on Computer Science and Education, ICCSE 2021*.
* Yavuz, E., & Atar, H. Y. (2020). An examination of Turkish students' PISA 2015 collaborative problem-solving competencies. *International Journal of Assessment Tools in Education, 7*(4), 588–606. https://doi.org/10.21449/ijate.682103.
* Yuan, J., Xiao, Y., & Liu, H. (2019). Assessment of collaborative problem solving based on process stream data: A new paradigm for extracting indicators and modeling dyad data. *Frontiers in Psychology, 10*, 369. https://doi.org/10.3389/fpsyg.2019.00369.
Zurita, G., Nussbaum, M., & Salinas, R. (2005). Dynamic grouping in collaborative learning supported by wireless handhelds. *Journal of Educational Technology & Society, 8*(3), 149–161.
Note: the articles synthesized in the literature review are designated by an asterisk.