An In-Depth Study of Assessment of Collaborative Problem Solving (CPS) Skills of Students in Both Technological and Authentic Learning Settings

Haley Wing Chi Tsang, Yiming Liu, and Nancy Wai Ying Law
haleywct@connect.hku.hk, eduliuym@connect.hku.hk, nlaw@hku.hk
The University of Hong Kong

Abstract: In a world dominated by ever-increasing technology-mediated social interactions over the now ubiquitous Internet, collaborative problem solving (CPS) has emerged as a critical 21st century skill for study, work and general well-being. In education, both learning and assessing CPS competency are imperatives. However, research related to CPS assessment has not been a mainstream topic in educational measurement, though there are international efforts like PISA and ATC21S initiatives, and a few assessment methodologies have been proposed. This study is an in-depth investigation to compare students' CPS performance in a closed-ended technology-mediated and an open-ended authentic task setting. Differences in assessment results were observed, which has important implications for the design of learning and/or assessment in CPS. In the extant literature, large-scale CPS assessments deploy tightly framed, technology-mediated tasks, while CSCL research studies learning in authentic contexts. Our findings contribute to a better understanding of how task contexts influence CPS behavior.

Keywords: collaborative problem solving, assessment, serious games, game design

Introduction

Collaborative problem solving (CPS) has become the dominant competency to be acquired by students in the 21st century (Fiore et al., 2017; Krkovic, Mustafic, Wüstenberg, & Greiff, 2018; OECD, 2017). CPS helps students seek better and quicker solutions in their studies. Compared to traditional problem solving skills, CPS necessitates a more diverse set of skills – cognitive as well as social – to learn and master before its potential benefits can be realized. This broader skill set is described in CPS conceptual or assessment frameworks. To date, notable CPS frameworks include the Assessment and Teaching of 21st Century Skills (ATC21S) (Hesse, Care, Buder, Sassenberg, & Griffin, 2015) and the Program for International Student Assessment (PISA) in education (OECD, 2017).

Getting to know the baseline information of how well a student performs in CPS is a prerequisite for effective teaching and learning of CPS. However, methods available for CPS assessment have not been many. The recent ones are computer-based, simulation-based or game-based assessment, as exemplified by the large-scale, international assessments program of PISA (OECD, 2017) and ATC21S (Griffin & Care, 2015) as well as smaller-scale assessment studies from Educational Testing Service (ETS) (e.g. Liu, Hao, von Davier, Kyllonen, & Zapata-Rivera, 2016, and von Davier, Hao, Liu, & Kyllonen, 2017) and other scholars (e.g. Rosen, 2017, and Yuan, Xiao, & Liu, 2019). These offer a virtual world which resembles or simulates real-world problem solving and collaboration scenarios and flexibly facilitate the CPS process as it unfolds. However, what students might do is to a certain degree bound by what is set up or made available by the assessment design right from the start. On the other hand, in authentic, open-ended CPS process commonly found in collaborative curriculum projects, students could interact more freely and demonstrate the requisite skills more naturally (Bause, Brich, Wesslein & Hesse, 2018), suggesting that assessment in such setting would also be necessary and beneficial.

The foregoing account has motivated the pursuit of answers to the following two research questions in this study:

a) Would there be discrepancies in CPS assessment results between measurements in both technology-mediated and open-ended, authentic task settings?
b) What insight may CPS assessment results shed on learning designs for CPS skills development?

Literature review

Collaborative learning and problem solving

Computer-supported collaborative learning (CSCL) studies the ways in which people can learn together with the support of computers and is a vision focusing on what technology can promise in collaborative learning and what related research should be conducted (Stahl, Koschmann, & Suthers, 2006). For more than twenty years, the field of CSCL has evolved and progressed in parallel with developments in ICT. In particular, mobile technology has already established a history of its presence in CSCL and more recently, educational video games have been used and studied in CSCL as well (Wake, Guribye, & Wasson, 2018). On the other hand, CPS refers to the process undertaken by a group of students during which they solve a problem together and create knowledge (Cukurova, Luckin, Millán, & Mavrikis, 2018). It has become an important part of CSCL (Hesse et al., 2015) as students solve problems to learn in collaborative learning and it therefore shares many important attributes and elements of CSCL. CPS is mainly concerned with creating and maintaining a shared, socially negotiated conceptual problem space where coordinated, synchronous activities of engagement and social interaction occur, and cognitive abilities of students are exercised in joint problem solving (Roschelle & Teasley, 1995). In particular, the reason for the rising importance of CPS is that today's problems are getting more complex and beyond one's capacity to solve them alone in many occasions – a common phenomenon in the 21st century. One solution is to solve them by a group of persons with different kinds of knowledge, skills and experiences (OECD, 2017). Therefore, many educational researchers, practitioners and policy makers regard that CPS competency is important to students and its acquisition in school is essential, preparing them for future study and work.

CPS skill assessment framework: ATC21S

Collaborative learning and CPS skills are not developed in vacuum and a meaningful, authentic context is required for them to flourish (Jonassen, 2004; Nicaise, Gibney, & Crane, 2000). To motivate students to practice and display either skill, a learning environment which stimulates creativity and challenges their cognitive abilities in a social, interactive setting would engage them and encourage them to display their maximum potential capabilities (Stahl et al., 2006). If they are doing creative tasks to solve a problem with members in the same group, the synergistic effects of collaboration will benefit the learning and practice of CPS. In the CPS/CSCL context, this would often require the development of new software applications (e.g. simulations and serious games) to foster an engaging virtual world for connecting group members socially to explore or perform creative tasks to seek solutions to what they do not know (Stahl et al., 2006). Creativity could be a vision in CSCL (Stahl et al., 2006), an embedded part in social groups (Sawyer & DeZutter, 2009), and an outcome of collaboration (Sawyer, 2017). At the same time, to help students get feedback from what they have done in the collaboration for improvement, motivation, and progress monitoring, it is essential that their skills exhibited are being assessed (Griffin & Care, 2015). However, CPS assessment has not been common both locally and internationally. At the international, large-scale level, there are two major frameworks in education developed for the assessment of CPS (Fiore et al., 2017): the ATC21S framework for the 11 to 15 years old students (Griffin & Care, 2015) and the PISA one for the 15 years old students (OECD, 2017). In this study of CPS skills of students, the ATC21S framework is adopted because it is a framework for teaching and assessing students' skills, its associated computer-based assessment tasks developed on the basis of the framework are used internationally, and its levels of complexity in theory, structure and practice are deemed reasonable and appropriate for the current investigation.

ATC21S is a comprehensive CPS assessment framework which divides the CPS skills into three levels (Hesse et al., 2015). At Level 1, there are (1) cognitive skills and (2) social skills. The first refers to the ability to reason and regulate fine details of a solving process from problem identification and definition, solving procedures, resource acquisition, monitoring to reflection while the second refers to the ability to perform inter-personal communication and interactions, and coordinate and direct activities in a social environment. The cognitive process skills are further divided into two Level 2 skills: (i) learning and knowledge building skills (ability to identify, understand and formulate relationships between different elements of a problem, and enhance representations as well as refine solutions), and (ii) task regulation skills (ability to coordinate and manage different aspects of the entire problem solving process such as problem analysis and resource management). On the other hand, the social process skills are further divided into three Level 2 skills: (i) participation skills (ability to take part socially in the joint activities of a group), (ii) perspective taking skills (ability to view, understand, and interpret materials presented by persons or a situation from different perspectives involving knowledge in multiple disciplines), and (iii) social regulation skills (ability to handle, integrate and take advantages of diversities of group members in backgrounds, views, knowledge, skills and experiences). Each of the five Level 2 skills are further sub-divided into two to six sub-skills (Level 3 skills), making up a total of 18 Level 3 skills as shown in Table 1.

Table 1: The three-level ATC21S CPS assessment framework (Hesse et al., 2015, p. 41-52)

Level 1	**Level 2**	**Level 3**
(1)	(i) Learning and knowledge	Hypothesis "what if…" (Reflection and monitoring)

Level 1	Level 2	Level 3
Cognitive skills	building skills	Relationships (Representations and formulations)
		Rules: “If…then”
	(ii) Task regulation skills	Flexibility and ambiguity
		Goal setting
		Information collection
		Problem analysis
		Resource management
		Systematicity
(2) Social skills	(i) Participation skills	Action
		Interaction
		Task completion/perseverance
	(ii) Perspective taking skills	Adaptive responsiveness
		Audience awareness (mutual modelling)
	(iii) Social regulation skills	Self-evaluation (Metamemory)
		Negotiation
		Responsibility initiative
		Transactive memory

CPS assessment methods

There are two approaches available in constructing technology-based CPS assessment: human-to-human and human-to-agent. The decision on which approach to adopt is based upon testing environment, group composition, testing medium, scoring rules design, implementation, and others (Fiore et al., 2017; Griffin, Care, & Hardling, 2015). In the human-to-human approach, students participating in the test use face-to-face or computer-mediated communication such as chat box in collaboration (which is used by ATC21S). It is a less standardized test environment and the challenge is how to identify, control and take into consideration of the effects of behaviors, responses and stimuli of other students on a participating student (Herborn, Stadler, Mustafić, & Greiff, 2018) when actual human behaviors during assessment are difficult to predict beforehand. In the human-to-agent approach which is adopted by PISA, however, the test is more standardized because the participating student interacts with a computer agent instead during collaborative problem solving (Herborn et al., 2018) and the behaviors of agents and various facets or factors of the testing environment are under pre-programmed logic set up at the design time.

For CPS assessment by a more conventional means, one method is the in-depth assessment of CPS in face-to-face, authentic scenarios which has remained no less important to date and is indispensable (Bause et al., 2018; Stahl et al., 2006). This approach, which is adopted in this study, can offer advantages such as refined information in detail about and new insights into a specific area under investigation (Stahl, 2017; Stahl et al., 2006). It is especially useful for an exploratory study in an emerging field like CPS assessment where relevant information is not widely available. In the same investigation, it can more flexibly and dynamically include individual assessment, group assessment or others and reveal detailed characteristics and fresh relationships between them, thus addressing the absence of the group feature found in PISA or ATC21S among other valid concerns such as authenticity and natural interactions. These concerns are important considerations in assessment. Take authenticity as an example, to make CPS assessment be more relevant and offer more assistance to students in learning CPS, a CPS skill test context is preferably authentic, meaning that the problem to solve and the problem context are of general interest and close to daily life, and no specific subject matter knowledge is required in reaching a solution (Care & Kim, 2018; Jonassen, 2004; Nicaise et al., 2000).

Setting and method of study

The study adopts the case study approach (Yin, 2003) to investigate how a group of students collaborate to solve problems related to cyberbullying learning and game design. It assesses CPS behaviors and performance of students in a serious game setting which is technology-based as well as an open-ended, authentic setting which is more conventional. In both settings, students communicate with each other face-to-face. The study was conducted in a 2-hour session for groups of primary school students and is divided into four parts as follows.

1. Learning course – studying: Students attend a short course to obtain an overview of the course and activities, and learn the basics of collaboration and how to improve and design a serious game.
2. CPS game play task – playing: Students play an online multi-player serious game and solve the problems

presented in three game quests, which are about helping a cyberbullying victim to recover quickly.

3. Game design task – designing: As a group, students first discuss and exchange views on perceptions of and experiences with the game played. Then they attempt to solve a design problem related to enhancing the played game by generating ideas and potential solutions together with the use of worksheets and storyboard which consists of a sequence of scenes with descriptions, narratives and drawings showing the game design, avatars and flow.
4. Focus group interview – reflecting: With the researcher acting as a facilitator, students identify and discuss how they might have collaborated and performed better in the group activities, express personal experiences, and suggest ideas for improving problem solving and collaboration skills.

The group activities were videotaped to obtain footage of students' communications and interactions. The discourse in the video recordings was transcribed verbatim. To assess the CPS skills of students, the utterances made by students during the collaboration were categorized and then coded according to a coding scheme which is specially developed based on the ATC21S CPS assessment framework (Hesse et al., 2015). The coding scheme consists of two categories of codes, cognitive skills and social skills, which are representative of behavioral indicators of exhibited CPS skills inferred from the utterances. The final coding scheme emerges through a number of iterations of refinement, meaning that after each round of reviewing the codes set against the contents of the transcripts, necessary adjustments or enhancements in the coding scheme are made. When the utterances are fully coded, the number of occurrences of a CPS skill indicator of a student, known as frequency of the concerned skill, is obtained. The CPS assessment is then based on the frequency data of students. In this study, the coding was done by the principal researcher first. A second coder coded 50% of the transcripts as well. An inter-rater reliability analysis was performed. The results showed that the Cohen's kappa for the cognitive skill dimension ($\kappa = 0.75$, $p < .0005$) and social skill dimension ($\kappa = 0.83$, $p < .0005$) were considered to be "good" and "very good" respectively (Altman, 1999). A sample of the coding results of utterances in the transcripts is shown in Table 2 (The utterances were originally in Chinese and they were translated into English.).

Table 2: Examples of utterance coding

Setting	Skill category	Mapped ATC21S skill	Utterance	Context
Technological (game play task)	Cognitive	Flexibility and ambiguity	Don' mess up…. we need to…walk around the whole city first.	Make a suggestion on how to find the trusted person more quickly.
	Cognitive	Systematicity	We should not choose different persons, don't forget. Choosing different persons, then we cannot finish the task. Walk together.	Tell others what to do in order to make progress in the game.
	Social	Task completion/ perseverance	We must be serious.	Ask others to be more attentive in completing the quest.
Authentic (game design task)	Cognitive	Resource management	You have to come up with design ideas!	Tell a member what to do next after completing the current work.
	Social	Self-evaluation (Metamemory)	My drawing is ugly. Ok, it is done	It is about storyboard drawing in quest design.
	Social	Responsibility initiative	Concentrate on your work, don't play anymore.	Remind other members to focus on game design.

Results and data analysis

The participants were 34 Grade 3 students from a local primary school and in the study were divided into groups of three to four members. One group with higher video recording quality was selected for study and their demographics are shown in Table 3. The data analysis consists of two parts: a) skill frequency analysis, and b) skill diversity analysis for the game play task and game design task. The former analyzes data which are about the number of times of a skill being performed by a student, and the latter analyzes data which are about the number of distinct skills having been performed by a student.

Table 3: Demographics of students in the group

Student ID	Gender	Age
PR25	Female	9
PR26	Female	8
PR27	Male	9
PR28	Male	10

a) Skill frequency analysis

Table 4 looks at the assessment data from the skill frequency perspective and shows the frequencies of cognitive and social skills of each student and the group for both the game play task and game design task. For the game play task, Student-PR28 was noticeably the top performer in cognitive skills and his frequency was one and a half times of the total of the other three students. The site observation and video confirmed that he dominated the group to a great extent. This was remarkable for him. Perhaps one reason is that he was the oldest, being one year older than the group average or he was a top gamer. Student-PR26 got the lowest frequency which was only one sixth of that of the top performer. Student-PR28 was the best student in social skill frequency which, however, was not as outstanding as his cognitive skills with his social skill frequency being accounted for about 32% of the group total. The social skill frequency distribution in the group was more even than the cognitive skill frequency. Student-PR25 became the lowest performer and her frequency (77) is only half of that of the top performer, Student-PR28 and far below the group average (121.75).

In the game design task, the dominance of Student-PR28 continued but in a less dramatic way. He was again first in both cognitive and social skill frequencies. The cognitive performances of the other three students were not significantly different from each other. The cognitive skill frequencies of Student-PR26 and Student-PR27 were the lowest but not far from the group average (18 and 18 vs. 22.75). For social skill frequency, the performances of Student-PR25 and Student-PR26 were very near the group average (71 and 69 vs. 69.75). The performance of the last student, Student-PR27, was the lowest but he still managed to get his share of the group total to be not far from one quarter. Like the cognitive skill frequency, the frequency distribution could be said to be fairly even among the four students.

Table 4: Cognitive and social skill frequencies of the students and group

Setting	Skill category	Assessment	Student				Group total	Group average
			PR25	PR26	PR27	PR28		
Technological (game play task)	Cognitive skills	Frequency	8	5	8	33	54	13.5
		% of group total	14.81	9.26	14.81	61.11	/	/
		Ranking	2	3	2	1	/	/
	Social skills	Frequency	77	130	124	156	487	121.75
		% of group total	15.81	26.69	25.46	32.03	/	/
		Ranking	4	2	3	1	/	/
Authentic (game design task)	Cognitive skills	Frequency	26	18	18	29	91	22.75
		% of group total	28.57	19.78	19.78	31.87	/	/
		Ranking	2	3	3	1	/	/
	Social skills	Frequency	71	69	53	86	279	69.75
		% of group total	25.45	24.73	19.00	30.82	/	/
		Ranking	2	3	4	1	/	/

b) Skill diversity analysis

From the skill diversity perspective, Table 5 shows the number of distinct cognitive and social skills exhibited by the individual students and group as a whole (Note: The "Group" column shows the number of distinct skills exhibited by the group as a whole.). For cognitive skills, Student-PR28 had the highest number ("7") in both tasks and his number was about 1.5 to 2 times of those of his peers, which should be considered to be significant. The numbers of each student did not vary significantly between the two tasks. The group numbers of distinct skills were slightly higher in the game play task ("9" vs. "7"). The second position in both tasks went to Student-PR25. For social skills, Student-PR28 obtained the highest number in the first task but obtained the lowest instead in the second task, which, however, was only slightly lower than the highest ("6" vs. "7"). In the two tasks, every student got considerably higher numbers in the second task except Student-PR28 whose number was lower. Overall for social skills, Student-PR28 was still the winner but only performed slightly better than Student-PR27 who occupied the second position. The numbers of distinct social skills were the same ("8") in both settings. The performance of Student-PR28 was worth special mention because when both social and cognitive skills of the two tasks were considered as a whole, he was the first and possessed nearly all the distinct skills of the whole group. This phenomenon might be partly explained by the fact that he was more than one year older than other members.

Table 5: Number of distinct skills exhibited by the individuals and group

Setting	Skill	PR25	PR26	PR27	PR28	Group
Technological (game play task)	No. of distinct cognitive skills	5	4	3	7	9
	No. of distinct social skills	4	4	5	7	8
Authentic (game design task)	No. of distinct cognitive skills	4	3	3	7	7
	No. of distinct social skills	7	6	7	6	8

Discussion and conclusion

In the data analysis, how the skill frequency and diversity change from student to student in the game play and game design tasks has been described. The interpretations of the nature of these changes will be discussed next, leading to the answering of the two research questions related to CPS assessment and CPS learning design.

With regard to the first research question, the cognitive and social behaviors of participants have often been the focal point of CPS investigations. Referring to Table 4, when the cognitive skill frequency data of individual students in the game play task and game design tasks were compared using three views of assessment data separately: frequency, % of group frequency total and ranking, the patterns of students' performance in the two tasks were not the same except that Student-PR28 was repeatedly the first. The same findings applied to the social skill frequency. Therefore, on the whole, it can be said that there are discrepancies in assessment results between the two tasks for both cognitive and social skills – an inference being supported by three different ways of examining the data.

In greater detail, it is found that the frequencies of social skills exhibited by individual students are in general higher in the game play task than those in the game design task. This might be expected because students not only solve problems but also play in a socially friendly serious game which offers remarkable features such as vivid colours, lively 3D graphics visualization and simulation, an interesting storyline, and playfulness (Phillips & Popović, 2012), which particularly appeal to young students like participants in this study. A great number of social interactions happen when they play to solve problems or solve problems to play. They feel engaged and are immersed in the problem solving process and context. The game provides them opportunities at various interaction points as the game unfolds to express themselves at moments of joy or excitements and when formulating gaming strategies from instinct or not, for example. On the other hand, in the open-ended, authentic task, the element of playfulness is mostly absent (though one may argue that students may derive interests from the intellectual challenges of game design) and more disciplined or involved problem solving takes over, requiring moments of better cognitive skills but not as many moments of social skills as in the other task. This means that the less structured environment as a result of open-ended, authentic characteristic of the game design task fosters or encourages more cognitive skill expression compared to the more confined game play task in which opportunities or flexibility to express cognitive skills are often confined by the game design, not the students. Although students communicate less, this does not necessarily mean that they are not as active as before. The whole situation results in less social behaviors but leads to more cognitive behaviors of students than the game play task.

Taking advantage of the assessment results in assessment design is natural and meaningful. Then would it be possible to take advantage of the same assessment findings in CPS learning design as well? This begs the second research question: What insight may CPS assessment results shed on learning designs for CPS skills development? In order to answer this research question better, the same assessment data set is analyzed from another perspective: skill diversity perspective which deals with the number of distinct skills, cognitive or social, exhibited by a student in CPS. In general, a capable collaborative problem solver will employ selected distinct but relevant skills a suitable number of times to solve a given problem with other team members. Referring to

Table 5, from the cognitive skill diversity perspective, when the numbers of distinct skills of each student are compared between the two tasks, the two patterns of numbers about students do not show striking differences. However, from the social skill diversity perspective, a similar comparison shows that students generally exhibited more distinct social skills in the game design task than the game play task. An attempt could be made to explain the skill diversity findings.

Normally it would be expected that an open-ended, authentic setting like the game design task might increase the opportunities provided to students to adopt more relevant varieties of cognitive skill to arrive at a solution because such setting imposes less restriction and students may more likely attempt to think of how to solve a problem in an effective way. Therefore, the cognitive skill diversity of the game design task would be greater than the game play task which is a confined technological setting with freedom allowed in solving a presented problem limited by the quest design to a certain degree. However, this did not happen. In fact, the diversities in cognitive skills exhibited by students in both tasks are not significantly different. One possible reason is that young students in the design task did not have the necessary knowledge or competency in cognitive skills to take advantage of the open-ended environment to increase the cognitive skill diversity beyond what they have been achieved in the previous task. On the other hand, they were successful in increasing the social skill diversity in the open-ended setting to a level higher than that in the more structured game play setting perhaps because using more distinct social skills is relatively easier and more natural to them compared to cognitive skills.

Combining the findings based on Table 4 and Table 5, how assessment results might inform learning design in CPS from the skill frequency and diversity perspectives becomes clearer. First, from the social skill frequency perspective, the technological setting exemplified by the serious game shows that the engagement, interactivity, immersion, game mechanics, dynamics, and aesthetics of serious games can create many natural opportunities for students to communicate and work together, which are conducive to display of social skills in CPS learning. Therefore, CPS learning design should attempt to incorporate serious game elements (Shih et al., 2010) when possible to increase the quality of collaboration. Second, since it is found that perhaps because of students' inadequacy in cognitive competency, they could not increase the diversity level of cognitive skills when taking advantage of open-ended, authentic setting in the game design task, stronger pedagogical support measures are preferably to be included in CPS learning design to foster adoption of new variety of cognitive skills. For example, if assessment shows that students are weak in setting goals (which might apply to some students in this study), learning scaffoldings which provide more background and problem context information may become one of the design criteria. Third, assessment results of students have indicated that they could act or respond more frequently using cognitive skills they have already possessed (the cognitive skill frequency perspective) in a CPS environment such as the game design task (which is open-ended and authentic) favoring openness or freedom and allowing for flexibility in approach. In the same environment, they could also exhibit more distinct social skills (the social skill diversity perspective). This means that less structured and less confined learning space should be adopted when appropriate in learning design to encourage them to take initiatives to make use of their existing skills or even untapped talents for better learning outcomes.

Extending the discussion, the potential impact of the findings in this single study covering two important and connected settings for CPS assessment could be significant in the longer term from the learning sciences research perspective. The study has demonstrated that there are differences in assessment results between the two settings while the assessment model in each has put forward valid arguments in its own right from conceptual and operationalization points of view. ATC21S (human-to-human, dyads, computer-mediated textual communication) (Griffin & Care, 2015) and PISA (human-to-agent, human-computer pairs, simulated textual communication) (OECD, 2017) adopt one setting but this study (human-to-human, face-to-face communication) adopts both. For wide applicability and deep understanding, the present CPS assessment research mostly in the technological setting might transition to the dual-setting approach to take advantage of their respective strengths. As the study has shown that assessment results would inform CPS learning design, reflecting the close relationship between both, going forward, the learning sciences researchers could take note of the disparities and make a similar transition – from the authentic setting mostly to the dual-setting approach – for the same reason. The two potential moves might converge or even overlap with one another, bridging and boosting the two fields.

References

Altman, D. G. (1999). *Practical statistics for medical research.* New York: Chapman & Hall/CRC Press.

Bause, I. M., Brich, I. R., Wesslein, A. K., & Hesse, F. W. (2018). Using technological functions on a multi-touch table and their affordances to counteract biases and foster collaborative problem solving. *International Journal of Computer-Supported Collaborative Learning, 13*(1), 7-33.

Care E., & Kim H. (2018). Assessment of twenty-first century skills: The issue of authenticity. In E. Care, P. Griffin, M. Wilson (Eds.), Assessment and teaching of 21st century skills. Educational

assessment in an information age (pp. 21-39). Cham: Springer.
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.
Fiore, S., Graesser, A., Greiff, S., Griffin, P., Gong, B., Kyllonen, P., . . . von Davier, A. (2017). *Collaborative problem solving: Considerations for the National Assessment of Educational Progress*. Washington, D. C: National Center for Education Statistics (NCES).
Griffin, P., & Care, E. (2015). *Assessment and teaching of 21st century skills: Methods and approach*. Dordrecht: Springer Science+Business.
Griffin, P., Care, E., & Hardling, S. (2015). Task characteristics and calibration. In P. Griffin & E. Care (Eds.), Assessment and teaching of 21st century skill (pp. 133-178). Dordrecht: Springer Science+Business.
Herborn, K., Stadler, M., Mustafić, M., & Greiff, S. (2018). The assessment of collaborative problem solving in PISA 2015: Can computer agents replace humans? *Computers in Human Behavior*, *104,* 105624
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 (pp. 37-56). Dordrecht: Springer Science+Business.
Jonassen, D. H. (2004). *Learning to solve problems: An instructional design guide.* San Francisco, CA: Pfeiffer/Jossey-Bass.
Krkovic K., Mustafic M., Wüstenberg S., Greiff S. (2018). Shifts in the assessment of problem solving. In E. Care, P. Griffin, M. Wilson (Eds.), Assessment and teaching of 21st century skills. Educational assessment in an information age (pp. 53-73). Cham: Springer.
Liu, L., Hao, J., von Davier, A. A., Kyllonen, P., & Zapata-Rivera, J. (2016). A tough nut to crack: Measuring collaborative problem solving. In Y. Rosen, S. Ferrara, & M. Mosharraf (Eds.), Handbook of research on technology tools for real-world skill development (pp. 344-359). Hershey, PA: IGI Global.
Nicaise, M., Gibney, T., & Crane, M. (2000). Toward an understanding of authentic learning: Student perceptions of an authentic classroom. *Journal of Science Education and Technology, 9*(1), 79-94.
OECD. (2017). *PISA 2015 results (Volume V): Collaborative problem solving*. Paris: PISA, OECD Publishing.
Phillips, V., & Popović, Z. (2012). More than child's play: Games have potential learning and assessment tools. *Phi Delta Kappan, 94*(2), 26–30.
Roschelle, J., & Teasley, S. D. (1995). The construction of shared knowledge in collaborative problem solving. In C. E. O'Malley (Eds.), Computer supported collaborative learning (pp. 69-97). Berlin: Springer-Verlag
Rosen, Y. (2017). Assessing Students in Human-to-Agent Settings to Inform Collaborative Problem-Solving Learning. *Journal of Educational Measurement, 54*(1), 36-53.
Sawyer, K. (2017). *Group genius: The creative power of collaboration*. New York, NY: Basic Books.
Sawyer, R. K., & DeZutter, S. (2009). Distributed creativity: How collective creations emerge from collaboration. *Psychology of Aesthetics, Creativity, and the Arts, 3*(2), 81.
Shih, J. L., Shih, B. J., Shih, C. C., Su, H. Y., & Chuang, C. W. (2010). The influence of collaboration styles to children's cognitive performance in digital problem-solving game "William Adventure": A comparative case study. *Computers & Education, 55*(3), 982-993.
Stahl, G. (2017). Group practices: a new way of viewing CSCL. *International Journal of Computer-Supported Collaborative Learning, 12*(1), 113-126.
Stahl, G., Koschmann, T., & Suthers, D. (2006). Computer-supported collaborative learning: An historical perspective. In R. K. Sawyer (Eds.), Cambridge Handbook of the Learning Sciences (pp. 409-426) , UK: Cambridge University Press.
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.
Wake, J. D., Guribye, F., & Wasson, B. (2018). Learning through collaborative design of location-based games. *International Journal of Computer-Supported Collaborative Learning*, *13*(2), 167-187.
Yin, R. K. (2003). *Case study research: design and methods* (3 ed. Vol. 5). Thousand Oaks, CA: Sage Publications.
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).

Acknowledgements

The authors wish to acknowledge that this work is funded by the Research Grants Council of the HKSAR Government, #T44-707/16/N, under the Theme Based Research Scheme.