Assessing Collaborative Problem Solving: What and How?

Haley W. C. Tsang, Seung Won Park, Laure L. Chen, and Nancy W. Y. Law
haleywct@connect.hku.hk, swonpark@hku.hk, cllaure@connect.hku.hk, nlaw@hku.hk
The University of Hong Kong

Abstract: With the increasing interests in and recognition of the importance of CPS competence as an educational outcome goal, there will be stronger demands for valid instruments and metrics for the measurement of Collaborative Problem Solving Skills (CPS) from policy makers to education practitioners and parents. The availability of valid instruments and norms for CPS achievement would also promote CSCL adoption within the wider education arena. This study explores two methods of assessing CPS: using the ATC21S online tasks and using qualitative discourse analysis of the same groups of students' transactions as they work together on an authentic learning task. By comparing the measurements arising from both methods, we raise issues and contradictions observed and suggest further research directions related to the measurement of CPS.

Introduction

Collaborative problem solving (CPS) has been identified as an important competence critical for life and work in the 21st century in the education literature (Fiore et al., 2017) and in education policy documents (e.g. Singapore Ministry of Education (2009)). This is similar to what is stated in the P21 Framework for 21st Century Learning which is developed with participation from educators, business leaders and other stakeholders (Partnership for 21st Century Skills, 2009).

Collaborative learning (including CSCL) have gained increasing prominence as a field of study in education, and the general consensus on the importance of CPS skills, methods and tools for assessing CPS are limited. There are two major, popularly known CPS assessment frameworks and systems in education: the one developed by the ATC21S Consortium (Griffin & Care, 2015; Hesse, Care, Buder, Sassenberg, & Griffin, 2015), and the one used in the assessment of CPS in the Program for International Student Assessment (PISA) 2015 (OECD, 2017). Besides these two studies which have been administered to large populations of students in countries around the world, there are also smaller scale studies, such as those conducted by ETS in the US (Liu, Hao, von Davier, Kyllonen, & Zapata-Rivera, 2016) and researchers (Lin, Hsiao, Chu, Chang & Chien, 2015) in Taiwan. While these reported studies differ in terms of the content focus of the problems, the collaboration arrangements, and the technology platforms used, all of these assess CPS as an individual's attribute. On the other hand, studies in CSCL generally focus on the interactions and joint activities among group members and the performance of groups and communities. Koschmann (2001, p. 19) argue that the concern of CSCL is "with the unfolding process of meaning making …, not so-called "learning outcomes"". Stahl (2014) further elaborated that studies of collaborative learning focusing on the outcomes of individuals misses the most important aspect and educational potential of CSCL, that the group's outcome is not equal to the aggregate of the individual outcomes, and that the group is the appropriate unit of analysis for understanding collaborative learning. Despite the above differences between CPS assessment and CSCL, this paper will present a less traditional CPS assessment study with findings which may indicate the direction of future interplay between the two domains.

Literature review

Hesse et al. (2015, p. 38) define CPS as "approaching a problem responsively by working together and exchanging ideas". CPS thus comprises two constructs, collaboration and problem solving, as exemplified in the framework described, which is grounded on research in the CSCL area. Collaborative problem solving is similar to individual problem solving, except that in CPS all the steps are directly observable as the group members need to communicate and negotiate their understanding during each step of the process. Hesse et al's (2015) framework is adopted by the ATC21S Consortium as the assessment framework for their CPS assessment system. The PISA 2015 assessment framework (OECD, 2017) is also grounded on the same theoretical underpinnings as elaborated in Hesse et al's (2015), but expressed in a 4x3 matrix, comprising four stages of problem solving (exploring and understanding, representing and formulating, planning and executing, and monitoring and reflecting) and three communication processes (establishing and maintaining shared understanding, taking appropriate action to solve the problem, and establishing and maintaining team organization). Since in this study, the ATC21S assessment system has been deployed as one of the methods to assess students' CPS skills, we will further elaborate on this below.

The ATC21S assessment framework

The ATC21S assessment framework has a three-level hierarchy of component skills (Hesse et al., 2015), as shown in Table 1. At the top level, CPS comprises social and cognitive process skills. The former is further differentiated into participation skills, perspective-taking skills and social regulation skills, while the latter is differentiated into task regulation skills, and learning and knowledge building skills.

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

Level 1	Level 2	Level 3
(1) Social process skills	(i) Participation skills	(a) Action
		(b) Interaction
		(c) Task completion/perseverance
	(ii) Perspective taking skills	(a) Adaptive responsiveness
		(b) Audience awareness (mutual modelling)
	(iii) Social regulation skills	(a) Negotiation
		(b) Self-evaluation (Metamemory)
		(c) Transactive memory
		(d) Responsibility initiative
(2) Cognitive process skills	(i) Task regulation skills	(a) Problem analysis
		(b) Goal setting
		(c) Resource management
		(d) Flexibility and ambiguity
		(e) Information collection
		(f) Systematicity
	(ii) Learning and knowledge building skills	(a) Relationships (Representations and formulations)
		(b) Rules: "If…then"
		(c) Hypothesis "what if…" (Reflection and monitoring)

At the third or lowest level, participation skills include: (a) action – skills in performing the actions needed for the solving process, (b) interaction – skills of carrying out interactions with others in a chain of alternating actions (upon others) and in response to other's actions, and (c) task completion/perseverance – the ability to persist in the task performance until the task completion.

Perspective taking skills include: (a) adaptive responsiveness – skills of responding to others' views or contributions and incorporating them into one's own thoughts and actions, and (b) audience awareness (mutual modelling) – skills of adapting one's own views, efforts or accomplishments from others' (or audiences') perspectives when presenting them so that they are more easily understood or appreciated.

Social regulation skills include: (a) negotiation – resolving differences in work procedures, views, targets, priorities, potential solutions, etc. through compromise, consensus or other means; (b) self-evaluation (metamemory skill) – knowing one's position, knowledge, strengths, weaknesses and other attributes; (c) transactive memory skills – knowing other collaborators in the same terms as in knowing oneself in metamemory skills; and (d) responsibility initiative – taking initiatives and responsibilities at different stages of the problem-solving process to achieve high-quality solutions through the joint efforts of the group.

Task regulation skills include: (a) problem analysis – dividing a problem into a set of more manageable tasks for solving the problem; (b) goal setting – setting clear, specific and achievable goals and sub-goals to direct and motivate the group's problem solving efforts, (c) resource management – obtaining, allocating, utilizing and assessing resources based on the available pool of human resources (group members' knowledge, skills and special capabilities) and other tangible resources such as equipment, tools and software applications, (d) flexibility and ambiguity – broadening the range of solution methods through problem representation, re-organization or new strategies, and determining acceptable tolerance of ambiguity and treating them as opportunities to further explore the problem space for alternative steps or strategies; (e) information collection – identifying and collecting information to solve a problem to meet the needs in terms of relevance, scope, detail, and time availability; and (f) systematicity – using planful, methodical approaches such as means-ends analysis and adopting results monitoring and reflection

Learning and knowledge building skills include: (a) representation and formulations – identifying relationships between elements in the problem space, and making representations and formulations of the inner connections among elements of tasks, (b) rules – ability to recognize contingencies and causal relationships as in making "if…then" rules, and (c) hypothesis – setting hypotheses by generalizing observations or parts of the solution and testing them through "what if…" inquiries, followed by monitoring and reflection of actions and

strategies.

Design of CPS assessment systems

All CPS assessments systems are computer-based, and these can be generalized into two approaches. In the first approach, students work online on a collaborative task, communicating through a chat box, as exemplified by the ATC21S system (Awwal, Griffin, & Scalise, 2015). The CPS system licensed by the University of Melbourne for use by the research team only supports dyads (as randomly- or pre- assigned pairs) collaborating on a bundle of about three to four tasks, which are to be completed within an hour. The system is designed on the basis of a jigsaw model of collaboration such that any one person would not have enough information to complete the task without help from the collaborator. Process and click stream data generated during the task interactions are recorded in time-stamped session logs, and the chat box interactions are recorded in the chat logs. A sophisticated system of coding, mapping, and scoring was developed for each of the indicators based on the CPS assessment framework (Adams et al., 2015). The system can generate reports on individual and class (i.e. everyone who participated in the assessment in the same session) performance.

In the second approach, the PISA 2015 CPS assessment, items were constructed according to the assessment matrix such that each item can be classified as targeting one of the 12 problem-solving-task-focus and collaboration-focus combinations. Students collaborate with different numbers of computer agents and playing different roles in the interactions, according to the specific collaboration skills required (OECD, 2017). The extent to which other team members are able to collaborate can then be precisely controlled in this system. Student performance is assessed through the specific responses they select from the choices offered.

Both the ATC21S and PISA 2015 assessed CPS as a generic skill, and subject matter knowledge of the student in relation to the problem contexts is assumed to not affect the students' assessed CPS performance. A team from the US Educational Testing Service (ETS) studied students' CPS skills in the context of knowledge restructuring and revision in the study of science topics (Liu et al., 2016). They adopted the PISA 2015 CPS assessment framework in the design of their system, and conducted two studies, the first with individual students working with two computer agents, and the second with pairs of students working with two computer agents. They found that students make bigger learning gains as measured by the accuracy of their responses when they were collaborating with a human peer. However, students' CPS outcomes were not reported.

Nature of problem and roles of team members

In the CSCL literature, knowledge building (Scardamalia, 2002) is often the most valued targeted learning outcome in the process. Thus, open-ended authentic problems closely related or meaningful to students' everyday life and requiring some sustained engagement are often selected as the targeted context for collaborative learning. This is very different from the problems used in the ATC21S, PISA 2015 or the ETS assessment systems.

Another challenge that CPS assessment systems have to confront is the roles taken up by different members in the team, beyond those captured in the collaboration aspects of the assessment framework. One particular area of concern is the role of leadership in collaborative teams, which is an important construct in the management and organizational psychology literature (Mercier, Higgins, & Da Costa, 2014). Leadership refers to the skill of taking responsibilities as well as initiatives in activities of a group to direct the group to achieve its goals (Miller, Sun, Wu, & Anderson, 2013). Leadership roles can be a result of being elected (by peers) or assigned (e.g. by teachers), but can also be emergent, that is, as a dynamic property of the situation, resulting from the social and task interactions (Gressick & Derry, 2010).

Mercier et al. (2014) further differentiated emergent leadership into two types based on the nature of the moves made by the participants, which can be focusing on organizing the group, or on advancing the intellectual aspects of the group effort. They define three types of organizational leadership moves: turn management, planning and organizing, and acknowledgement; and two types of intellectual leadership moves: idea management and development, and topic control. Mercier et al (2014) further defined emergent leadership as instances when a participant's leadership moves were taken up by others in the group. Their study included findings that have important implications for how to design CPS assessment. First, they found that the extent to which students are more willing to work on ideas from other students is affected by the collaborative medium – being greater in electronic medium like a multi-touch tablet than the paper medium, which indicates that the former is more supportive of complex conversations. Second, they also found that more than one person exhibiting emergent leadership can occur in a group. Third, by studying emergent leadership when the same groups of students engaged in tasks associated with two different subject areas, math and history, they found that the emergence of leadership differed over different task types.

Study purpose and design

The study reported here is part of a bigger study that aims to develop an online role play serious games system that would support the learning and assessment of CPS for children aged 11 to 15. A key challenge in designing such a system is to develop valid measures for CPS assessment. In this study, our goal is to compare the CPS achievement of students as assessed by the ATC21S CPS system with a hand-coded CPS assessment based on video-recorded transactions of students working in groups over the space of two weeks to design an online game to promote cyberwellness. Their collaborative interactions were videotaped twice, with a week's separation in between. The specific research questions addressed in this study are as follows:

1. Is the ATC21S assessment framework appropriate for coding the face-to-face transactions of the students participating in the study?
2. How do students' CPS behavior and performance change over time, if at all?
3. Do the students' CPS achievement as measured by the ATC21S assessment system correlate with the hand-coded CPS scores?
4. What role can ATC21S CPS assessment scores play in the further development of CPS systems?

Settings and participants

The study was conducted in the context of a three-session summer course for primary and secondary school students on game design to promote cyberwellness. The game design course was offered to support students interested in participating in a Minecraft coding competition. Interested students registered through the online portal of the competition organizer, which did not set any selection mechanism or criteria on the applications for participation. The summer course was held as 2-hour workshop sessions on three consecutive Friday mornings. Students were grouped into teams of three to five students of approximately the same age. The goal of the course was for the students to develop jointly a game (storyboard) that would be both interesting and help others to learn about cyberbullying and how to handle them. The students were encouraged to build their games in Minecraft, but that was not a necessary part of the course activities. A brief rundown for the three workshop sessions is as follows:

Workshop 1: Welcome, introduction to the goals of the course, introduction to examples of and ways to handle cyberbullying, group assignment, and working in groups to discuss the goals of the game they wish to construct.

Workshop 2: Introduction to game design and development, including showing examples of storyboards and Minecraft games, and group work on game design: game goal, rules, and storyboard.

Workshop 3: Group preparation on presentation of preliminary game design to other groups for peer review and feedback, further refinement and submission of final game design storyboard, and group interviews.

A total of 44 students were admitted to the workshops on a first-come-first served basis without any selection criteria. According to the University of Melbourne Assessment Research Centre, the CPS tasks have only been trialed and validated for students aged between 11 and 15, and are deemed to be too difficult for younger students. Of the 44 course participants, only 29 were aged 11 or above. They were invited to arrive at the course venue an hour before the start of Workshop 2 to take the ATC21S CPS assessment. Of these, only 14 took the ATC21S CPS assessment and gave consent to the research team to use all of the artefacts they produced during the workshops as well as all audio and video recordings for research purposes.

The group work sessions in Workshops 2 and 3 were video-recorded. Workshop 1 was not videoed as the time period available for the group work was too short. The total time of the video recording for each group was around 55 minutes in Workshop 2, and 65 minutes in Workshop 3. For the purpose of this paper, we report on one of the groups with children aged 11 or above. All four members in this group attended the three workshops and took the ATC21S CPS assessment, and the video recordings from both workshop sessions were clear.

Results

The background of this group of students and their CPS scores in the ATC21S assessment are listed in Table 2.

Table 2: The students' demographic information and CPS achievement as assessed by the ATC21S CPS system

Student ID	Demographic information	Cognitive score	Cognitive level	Social score	Social level
GD28	Female, 14 year old, 8th grade	-0.46	3	0.85	5
GD29	Female, 11 year old, 6th grade	-0.33	3	0.94	5

| GD30 | Male, 12 year old, 6th grade | -0.45 | 3 | 0.89 | 5 |
| GD31 | Male, 12 year old, 6th grade | -0.19 | 3 | 1.84 | 6 |

Based on the ATC21S CPS assessment framework presented in Table 1 above, two of the research team members worked in parallel and then together to develop a detailed coding scheme for coding the video transcript. It is found that other than the off-task communications, all of the verbal transactions can be coded using the assessment framework. Other than the verbal transactions, the transcript also records instances where a member *starts* engaging in a task. These observation statements are coded under participation skills-action. An inter-rater reliability analysis on a sample of the dataset (25%) is performed to determine if there is agreement between the two independent coders. The Cohen's kappa ($\kappa = .82$, $p < .001$) obtained is considered to be "very good" (Alman, 1999).

As the present study is exploratory in nature, we decide to use a simple metric for our computation of CPS: that of frequency counts of turns made by each of the participants under the social and cognitive process skills categories. However, it is observed that while most of the discourse turns can be classified as instances of positive behavior in the subcategories, some of the turns are negative instances. For example, utterances such as "this is too difficult" or "I didn't say I want to do it, I am just saying …" are coded as '*avoids undertaking the work*' and is given a frequency count of "-1" to indicate that this is a negative instance of the code. Turns that are examples of ignoring other's ideas or arguing without giving reasons are considered negative instances of social regulation.

Table 3: The CPS scores computed based on the number of turns made by each student in the two workshops

	Workshop 2				Workshop 3				Overall			
Student ID	GD 28	GD 29	GD 30	GD 31	GD 28	GD 29	GD 30	GD 31	GD 28	GD 29	GD 30	GD 31
Social (+ve)	39	34	37	24	35	26	38	41	74	60	75	65
Cognitive (+ve)	22	9	10	9	8	9	16	11	30	18	26	20
Social (-ve)	7	12	5	7	3	6	5	9	10	18	10	16
Cognitive (-ve)	2	3	5	0	3	0	7	0	5	3	12	0
Social score	**32**	**22**	**32**	**17**	**32**	**20**	**33**	**32**	**64**	**42**	**65**	**49**
Cognitive score	**20**	**6**	**5**	**9**	**5**	**9**	**9**	**11**	**25**	**15**	**14**	**20**

The scores in Table 3 show a significant change in the scores associated with some of the students over the two workshops, both in terms of absolute numbers and relative to each other. The scores for GD29 and GD30 were relatively stable across the two workshops, and both increased in their cognitive scores. However, it is notable that GD29 has greatly reduced the frequencies of negative contributions in both social (from 12 to 6) and cognitive (from 3 to 0) dimensions. For GD28, while she maintained a high social score between the two workshops, her cognitive scores appeared to have been greatly reduced. On the other hand, in Workshop 3, GD31 increased significantly in his social score to rank the same with GD28 and GD30, and his cognitive score also rose to be the top score of the team. If we look at the overall scores over the two workshops, then GD28 and GD30 has the highest social score while GD28 has the highest cognitive score.

Comparing the scores in Tables 2 and 3, there is consistency in that the social scores of the students were much higher than their cognitive score. In fact, according to the ATC21S scoring rubric, 6 is the highest level of performance that a student can achieve. However, according to the scores in Table 2, GD31 has the highest score in both the social and cognitive dimensions while GD28 has the lowest scores in both the social and cognitive dimensions. How can we interpret these scores? Do they tell us anything about the students' CPS ability at all?

Student's perception of their own collaboration process

Towards the end of Workshop 3, a focus group interview was held with each of the participating groups. The following were the interview questions regarding how the team worked together to develop the game:

1. How did the team come up with the game ideas presented in their storyboard?
2. How did the team go about their division of labor for the project?
3. Did the team have to look for information to complete their project, and if so, what and how?
4. Did any contradictory views arise during the process, and if so, how were these resolved?
5. Did the team encounter any problems or challenges during the work process?
6. If you are given a second chance to create a game, will you do things differently?

While these questions were designed to find out the members' perceptions about various aspects of the

collaboration process, a surprising finding was that the role of GD28 as a leader was talked about by all the other students throughout the interview. The following is an excerpt from the part of the interview about division of labor:

GD30:   *We (GD29, GD31, GD30) argued when deciding whom to be the presenter.*

GD28:   *Exactly. Every time I was forced to present.*

...

GD30:   *Then, we would tell her (GD28) the key points and she will be the one who present them. That's all.*

GD31:   *As an analogy, we (GD29, GD31, GD30) are like the sheep while you (GD28) are the shepherd.*

Again, when asked how they would work differently if they were to create a game again (I stands for the Interviewer):

GD28:   *I will run away.*

GD29:   *The first thing I will do is…*

GD31:   *First, I will get the team leader back. Then…*

GD29:   *First I will get the team leader back and lock her up. Then...*

I: *That means you all want to listen to her instructions.*

GD30:   *Yes.*

It is clear from these exchanges that GD29, GD30 and GD31 unanimously consider GD28 as their leader in the project. According to GD30, GD28 played two roles—resolving conflict and representing the team:

GD30:   *The three of us often had different opinions. We would nominate our team leader whenever there was a presentation.*

Observation of the students' collaboration process

While rigorous coding of the students' turn-by-turn transactions provide us with a theoretically grounded view of the collaboration process, it is important that we do have a qualitative, holistic sense of how the students actually worked and interacted during the two workshop sessions. It should be noted that of the four students, GD30 was the only one who did not know how to program in Minecraft. Below is a brief summary of the key episodes:

Workshop 2 (Students were asked to work on their game design following the distributed worksheets)

1. *Deciding on the group name: GD28 led the discussion*. GD28 and GD30 constructed one using GD31's name, amid protests from GD31.

2. *Discussing game ideas*. GD28 suggested popping up questions for players to answer in Minecraft as the game mechanics. GD31 was more interested in playing around with Minecraft than completing the worksheets. GD30 requested GD31 to put aside the laptop to focus on the worksheet. GD28 then prompted the group for more game interaction ideas. GD30 put forward his ideas after hearing examples from GD28.

3. *Working on storyboard—division of labor*. GD29 asked what she should draw. After GD28 made her suggestion, GD29 expressed her desire to work in Minecraft instead. With help from GD30, GD29 was finally persuaded to work on the drawing task. GD30 then assigned a task to GD31 without giving reason, thus starting a conflict with GD31.

4. *Discussing the detailed design of the game quests*. GD28 initiated the discussion. GD29 and GD31 participated actively, but GD30 interrupted to say that this should be left to GD28 as the decision maker. GD28 paid close attention to GD29 and GD31's ideas. GD30 then join in to give his ideas.

5. *Starting to draw the storyboard*. GD28 gave suggestions to GD29 and GD31 on how to build the game in Minecraft. GD30 did not take up specific jobs but assigned jobs to GD29 and GD31.

Workshop 3 (Students were asked to give a short presentation on their game design as completed in Workshop 2 for peer feedback, and were given some time to prepare for it. After this, students were asked to work further and complete everything for a final presentation.)

1. *Discussing who should be the presenter*. GD29 and GD31 was playing on a smartphone. GD30 urged them to focus on the group discussion.

2. *Discussing the group name*. GD31 wanted to change, but no one else agreed.

3. *Discussing the detailed design of the game quests*. GD29 initiated the discussion and put forward her idea. GD30 built on her ideas and made further suggestions. GD31 disagreed with GD30 and both argued

strongly. GD28 stepped in to make other suggestions. A final agreement was reached.

4. *Creating the game in Minecraft.* GD31 expressed doubts about GD29's competence in Minecraft, ignored GD29's suggestions, turned the computer towards himself, and said that GD29 was stupid and was doing useless work. GD30 told GD31 that he was very impolite. GD29 and GD31 then entered into an intense argument over different approaches to work in Minecraft.

5. *Preparing for the first presentation.* GD28 took up the presenter role as she was considered the leader by the others. she practiced for the group presentation. GD30 asked how the game would handle the case if the player fail to achieve the game goal. GD28 and GD30 discussed the game refinement.

6. *First presentation.* Only GD28 came out to presented on behalf of the group. Not much feedback.

7. *Continuing the storyboard work.* GD28 assigned tasks to GD29 and GD31, GD30 worked with GD28 to help the other two students to complete the assigned tasks. GD29 and GD31 continued to argue, GD30 helped to clarify what needed to be done, and also helped GD31 with the task completion.

8. *Preparing for the group presentation.* GD31 finished creating part of the game in Minecraft. GD29 expressed delight and appreciation of the work by GD31. GD30 suggested to present not only the required paper-based storyboard, but also the partially completed game in Minecraft. GD31 referred to himself as the technical support for the group in building the game in Minecraft.

9. *Finalizing the game storyboards.* GD28 worked on finishing the writing part of the game storyboards. GD29 and GD30 worked together to stick pieces of the game scene paper drawings on the big cardboard for presentation.

10. Final presentation. GD28 was the main presenter, accompanied by the entire team. The Minecraft game was presented.

It can be seen that in Workshop 3, GD28 talked much less as she was focused on preparing for the two presentations. During this time GD30 was helping as the person to regulate the group and to resolve conflicts.

Discussion and conclusion

We begin our discussion by addressing the four research questions of this study. First, the ATC21S assessment framework was operationalized in the hand-coding of the CPS interactions, and found to be adequate for analyzing CPS in the face-to-face context. Second, as revealed by both the CPS scores and the observations of the collaboration process, the CPS behavior and performance of individual students and the team as a whole changed over time. Third, no conclusive correlation is observed between the ATC21S assessment scores and the hand-coded scores, whether for individual workshops or aggregated over both workshops. The fourth question about how the ATC21S assessment scores may contribute to future development of CPS assessment systems is more complex, and is discussed below in terms of the issues we have uncovered in this study.

As shown in our study, even when the same assessment framework is used for the assessment, the exhibited CPS behavior (on which the score is computed) is a fluid measure, dependent on the problem context and the stage of the problem-solving task progression. GD28 may seem to have declined in her leadership role in Workshop 3 based on the scores. She had in fact been engaging much more in the completion of the worksheet task. As the group settled into a more task focused mode of operation, she did not have to spend as much time or effort on regulating the group's activities. Moreover, we can see that GD30 had been assisting her in the coordination of the project. GD30 might have been able to engage more in the task completion part of the teamwork if he had knowledge of programming in Minecraft, and may not be doing so much task and social regulation work. The particular way a student engages in a task is always dependent on the nature of the task, the familiarity of the team members with the content/disciplinary knowledge and skills required in solving the problem, their familiarity with each other, and the stage of the problem solving progress. Is it meaningful to assess the CPS skills of individuals? What might be meaningful measures of a team's CPS skills? What contextual information might need to be specified for meaningful interpretation and use of CPS measures?

From our observations of the team's work progress, the conflicts arose mainly from the need to agree on what to enter on the worksheet and how to create the game in Minecraft. Through the interactions and conflict resolution process, the students improved in their teamwork to become more productive in their collaboration. However, the scores of the team members in Workshops 2 and 3 are not able to reveal the fine grain differences and development. These scores, similar to those from the ATC21S and PISA 2015, focus only on the CPS process but not the outcomes. The primary reason for valuing CPS as a core competence is the assumption that those with higher CPS capability will delivered better, higher quality collaboration outcomes. Is it reasonable to measure CPS capability, whether as an individual or group attribute, without giving any regard to the quality of the products of the collaboration? If CPS outcome were to be assessed, how should this be done?

The four students' CPS scores have been computed by the ATC21S system, which has been developed on the basis of an enormous amount of data and shown to be reliable for the set of assessment tasks in the system. However, there is as yet no reported studies on the construct validity of these measures, i.e. how these CPS scores might inform us on the assessee's CPS ability in authentic collaboration and task settings. What might be educationally meaningful ways of using CPS assessment scores?

Our study shows that there is a strong need for more studies on assessing CPS. So far, efforts to assess CPS have been led by researchers in the psychometric and assessment communities, focusing on CPS as an individual attribute. Many of the researchers in the CSCL community see CPS as a group attribute, and outcomes are often reported within the specific study context, but not in terms of a generic individual or group attribute. We hope the issues we have uncovered will stimulate interactions between the CSCL and assessment communities.

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* (pp. 115-132). Dordrecht: Springer.

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

Awwal, N., Griffin, P., & Scalise, S. (2015). Platforms for delivery of collaborative tasks. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 105-113). Dordrecht: Springer.

Fiore, S. M., Graesser, A., Greiff, S., Griffin, P., Gong, B., Kyllonen, P., . . . Rothman, R. (2017). *Collaborative problem solving: considerations for the National Assessment of Educational Progress.* Washington DC: National Center for Education Statistics (NCES).

Gressick, J., & Derry, S. J. (2010). Distributed leadership in online groups. *International Journal of Computer-Supported Collaborative Learning, 5*(2), 211-236.

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

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.

Koschmann, T. (2001, December). Revisiting the paradigms of instructional technology. In *Meeting at the crossroads. Proceedings of the 18th Annual Conference of the Australian Society for Computers in Learning in Tertiary Education* (15-22).

Lin, K. Y., Yu, K. C., Hsiao, H. S., Chu, Y. H., Chang, Y. S., & Chien, Y. H. (2015). Design of an assessment system for collaborative problem solving in STEM education. *Journal of Computers in Education, 2*(3), 301-322.

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

Mercier, E. M., Higgins, S. E., & Da Costa, L. (2014). Different leaders: Emergent organizational and intellectual leadership in children's collaborative learning groups. *International Journal of Computer-Supported Collaborative Learning, 9*(4), 397-432.

Miller, B., Sun, J., Wu, X., & Anderson, R. C. (2013). Child leaders in collaborative groups. In C. E. Hmelo-Silver, C. A. Chinn, C. Chan & A. M. O'Donnell (Eds.), *The international handbook of collaborative learning* (pp. 268-279). Routledge.

OECD. (2017). *PISA 2015 Results collaborative problem-solving.* Paris: OECD.

Partnership for 21st Century Skills. (2009). P21 framework definitions. Retrieved from http://www.p21.org/storage/documents/P21_Framework_Definitions.pdf

Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Eds.), Liberal Education in a Knowledge Society (pp. 67-98). Chicago: Open Court.

Singapore Ministry of Education. (2009). Third masterplan for IT in education (2009-2014). Retrieved from https://ictconnection.moe.edu.sg/masterplan-4/our-ict-journey/masterplan-3

Stahl, G. (2014). The group as paradigmatic unit of analysis: The contested relationship of CSCL to the learning sciences. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.644.8856&rep=rep1&type=pdf

Acknowledgements