



Collaborative Problem-Solving Behavior of 15-Year-Old Taiwanese Students in Science Education

Cheng-Hsuan Li

Graduate Institute of Educational Information and Measurement, National Taichung University of Education, TAIWAN

Zhi-Yong Liu

Graduate Institute of Educational Information and Measurement, National Taichung University of Education, TAIWAN

Received 8 June 2017 • Revised 25 August 2017 • Accepted 14 September 2017

ABSTRACT

Collaborative problem solving (CPS) is a crucial skill for students and people entering the workforce. Hence, an online CPS assessment system was developed for the Programme for International Student Assessment to test 12 CPS skills. This study compared the CPS skills of students on the basis of a CPS assessment developed for science scenarios in Taiwan. Moreover, a sequential analysis was applied to explore the behavioral patterns of students with CPS skills. The results demonstrated that most Taiwanese students are proficient in the 11 CPS skills (except for D3: “Monitoring, providing feedback and adapting the team organisation and roles”), and that female students are significantly more proficient than male students. The results also indicated that most students can successfully transition between CPS skills. However, the students in a given class may demonstrate a lack of behavioral patterns. Teachers can design activities and develop strategies to address this lack of behavioral patterns and thus increase students’ CPS skills and behavioral patterns.

Keywords: assessment, behavior pattern, collaboration problem solving, science education

INTRODUCTION

In the 21st century, noncollaborative work by a single person may be insufficient for solving complex problems, and it is generally preferable for people to solve problems together through collaboration and communication with other team members (Griffin et al., 2011; National Research Council, 2011). Hence, the teaching and assessment of students’ collaborative problem solving (CPS) skills is crucial to their preparation for future careers (Brannick & Prince, 1997; Griffin et al., 2011; National Research Council, 2011; Rosen & Rimor, 2012). Various human-to-human CPS tasks were developed to assess multidimensional skills, including social skills (participation, perspective-taking, and social regulation skills) and cognitive skills (planning, executing and monitoring, flexibility, and learning skills) for the Assessment & Teaching of 21st Century Skills (ATC21S) project. A person can collaborate with a partner to solve common problems online. Moreover, an assessment system can record corresponding conversations and actions in a log file (Griffin et al., 2011). Nevertheless, the level of individual skill should be graded by experts through rubrics. In addition, individual skill level may be influenced through collaboration with partners.

Therefore, the Organisation for Economic Co-operation and Development (OECD) designed a human-to-agent, computer-based standardization CPS assessment system for evaluating the CPS proficiency of 15-year-old students in the Programme for International Student Assessment (PISA) 2015 (OECD, 2013). The CPS skills were

© **Authors.** Terms and conditions of Creative Commons Attribution 4.0 International (CC BY 4.0) apply.

Correspondence: Cheng-Hsuan Li, *Graduate Institute of Educational Information and Measurement, National Taichung University of Education, Taiwan.*

✉ chenghsuanli@gmail.com

Contribution of this paper to the literature

- In this study, the levels of mastering of Taiwanese’s students for the 11 CPS skills, four problem-solving processes, and three major CPS competencies were analyzed according to two simulated CPS assessments in science problem scenarios.
- A coding scheme was proposed for conversation-dialog tasks that were assessed using predefined multiple choice test items. In addition, sequence analysis was applied to students CPS behavior in the science problem scenarios.
- The main contribution of the current study was to propose an analysis of students’ CPS behavioral patterns. Teachers can understand students’ CPS behavior patterns through the proposed coding scheme with other CPS assessment units.

based on other CPS frameworks, such as the Center for Research on Evaluation, Standards and Student Testing teamwork processing model (O’Neil et al., 2003, 2010), the teamwork model of Salas et al. (Fiore et al., 2008, 2010; Salas et al., 1992, 2008), and the ATC21S (Griffin et al., 2011), and it incorporated four cognitive processes into the individual problem-solving skills proposed in the PISA 2012 with three major CPS competencies (OECD, 2010; OECD, 2013).

Table 1 provides a matrix of CPS skills for the PISA 2015. The first column indicates the four major individual problem-solving processes: exploring and understanding, representing and formulating, planning and executing, and monitoring and reflecting (OECD, 2010, pp. 20–21). The first row displays the three major CPS competencies: establishing and maintaining a shared understanding, engaging in the appropriate action to solve a problem, and establishing and maintaining team organization (OECD, 2013). Each entry shows the interaction between the individual problem-solving process and CPS competency. Moreover, two sample units, the Aquarium (a test taker and a computer agent) and Class Logo (a test taker and two computer agents), were provided to assist readers in understanding the concepts of the assessment framework (OECD, 2013).

During a collaboration, an individual must communicate with and react to other participants to perform actions to achieve common goals. Moreover, individuals must sense and adapt to change their environment, and learn from each other (Franklin & Graesser, 1996; OECD, 2013). Because interactions between two or more participants are required for the CPS units, at least one other (computer) agent is required to collaborate with the test taker. In complex social interactions, some individuals support and praise other team members, whereas other individuals disagree with team members and provide negative feedback on the basis of false evidence. Furthermore, participants may have a different statuses, perspectives, abilities, and understandings when attempting to solve the problem (Millis et al., 2011; Wiley & Jensen, 2007). Therefore, single and multiple computer

Table 1. Matrix of Collaborative Problem Solving Skills for the PISA 2015

	(1) Establishing and maintaining shared understanding	(2) Taking appropriate action to solve the problem	(3) Establishing and maintaining team organisation
(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 organisation (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 organisation and roles

agent units were designed for the PISA 2015; the CPS skills shown in **Table 1** were assessed on the basis of interactions between an individual and computer agent(s) (OECD, 2013). In addition, units with a single computer agent and those with two computer agents were considered in this study.

Because students' domain knowledge influences collaboration and problem-solving processes, CPS is typically taught separately domains such as science, mathematics, and history (Funke & Frensch, 2007; Healy et al., 2002; Lee & Pennington, 1993; Mayer, 1992; Mayer & Wittrock, 1996; OECD, 2013). The PISA 2015 CPS assessment included domains such as mathematics, science, reading, environmental studies, community studies, and politics (OECD, 2013). In addition, science education is a crucial for young learners and is an obligatory element of the school curriculum. Furthermore, scientific literacy is a major domain in the PISA (OECD, 2003, 2013). During the communication and knowledge sharing that constitute CPS, students can explain the problem or their understanding of it to other team members and synthesize their understanding with knowledge from other subject areas (Luckin, Baines, Cukurova, & Holmes, 2017). Hence, they have opportunities for scientific argumentation including proposing, supporting, criticizing, evaluating, and refining ideas, and, thus, advance their beliefs about scientific phenomena. This emphasizes the usefulness of explanation and concept-mapping techniques as evaluative measures of student knowledge (Coleman, 1998; Duschl & Osborne, 2002). Some studies have revealed that students can work more collaboratively in science than in mathematics and humanities (Baines, Blatchford, & Kutnick, 2003; Kutnick & Blatchford, 2013; Webb & Palincsar, 1996; Luckin et al., 2017). Hence, this paper only discusses student CPS skill proficiency and behavioral patterns in the context of science.

On the basis of the PISA 2015 draft CPS framework and the corresponding 15 CPS skills, a CPS assessment system was developed for Taiwanese students in Grades 9 and 10 (aged approximately 15 years; Kuo, 2014; Li, Pai, Kuo, Lin, & Liu, 2015). The assessment included six units, two of which involved science-based scenarios (i.e., the Slurpee and Water Purification). The design concept of the Slurpee scenario was similar to that of the Aquarium sample unit (a test taker and a computer agent) in the draft CPS framework. In this case, the test taker and a computer agent collaborated to determine the optimal proportions of a cryogen. The design concept of the Water Purification scenario was similar to that of the Class Logo sample unit (a test taker and two computer agents) in the draft. The test taker and two computer agents collaborated to design an optimal purification bottle filter using five provided tools. The test taker was able to interact with one or more computer agents (simulated team members) through a chatbox to obtain the solution to a given problem from multiple choice items (OECD, 2013, 2016). Each item reflected one of the target skills in **Table 1**. The alternatives for an item were a set of chat options designed by experts; the test taker was instructed to select the most appropriate alternative. This study determined the skill level in Taiwanese students and compared skill differences with respect to the background variable of gender.

These two units were designed according to the PISA 2015 draft CPS framework (Kuo, 2014; OECD, 2013). Every task item corresponded to a target CPS skill with possible scores of 0, 1, and 2, which were predetermined according to chatting paths of the item. That is, the levels of mastery of the 12 CPS skills were scored automatically by the system. Hence, for each unit, the target CPS skills formed a fixed sequence and indicated possible skill changes among the 12 CPS skills. For example, if an item corresponded to matrix component C1 ("Communicating with team members about the actions to be/being performed"), and the next item corresponded to C2 ("Enacting plans"), then according to the graded skill levels, the item would determine whether students can enact plans after communicating with team members. Progressive sequential analysis can be used to infer potential behavioral patterns in CPS assessment (Hou, 2010). Furthermore, sequential analysis can indicate whether a certain behavior is followed by another through a statistical dependence index, which illustrates the difference between the observed behavioral pattern frequency and expected independent behavioral pattern frequency (Bakeman & Quera, 1995; Cheng & Hou, 2015; Hou, Chang, & Sung, 2007, 2008, 2009). However, during the assessment, the target CPS skill corresponding to an item had three levels, namely 0, 1, and 2. The combinations of these levels indicated variation among skill components. For example, scores of 2 on C1 and C2 indicates that the test taker can enact plans after communicating effectively; scores of 2 on C1 and 0 on C2 demonstrate that the test taker cannot enact plans efficiently. Therefore, a coding scheme for the CPS assessment was proposed in this study, and the differences between student behaviors were explored by applying sequential analysis to the CPS skill sequence on the basis of the proposed coding scheme.

In an academic achievement, educational researchers pay attention to gender gaps, especially in reading, mathematics, and science literacy (Ma, 2008). In science, boys have outperformed girls in the past, and gender differences have been observed in their interests and expectations from an early age, which have been attributed to the differences in their thinking, learning, and working styles (Browne & Ross, 1991; Murphy, 1997; Ziegler & Heller, 1997; Kimbell, Stables, Wheeler, Wosniak, & Kelly, 1991). Moreover, girls tend to work through communication and collaboration, whereas boys tend to work independently (Kimbell et al., 1991). However, studies on the gender gaps in science education have raised concerns about the issue of equity in science education; thus, gender differences in science have exhibited a general decrease since the 1980s. Furthermore, they have the same positive attitudes and interests in science. (Ziegler & Heller, 1997; Yip, Chiu, & Ho, 2004; Ma, 2008; Sjøberg & Schreiner, 2005). Nevertheless, collaborative style also affects science achievement among students (Schroeder, Scott, Tolson, Huang, & Lee, 2007). Therefore, this study also investigated the gender gaps in the participants' CPS skill proficiency and their behavioral patterns on the basis of the science scenarios.

This study explored CPS skill proficiency according to the responses to the CPS assessment developed by Kuo (2014; Li et al. 2015). Additionally, the sequential behavioral patterns of students with various CPS skill levels were compared. The research questions explored in this study were as follows:

1. What are the 12 CPS skill proficiencies for Taiwanese students in science scenarios?
2. Do the CPS skill proficiencies of male and female students differ significantly?
3. Which behavioral patterns do Taiwanese students lack in science scenarios?
4. Which behavioral patterns do the students in a particular class lack?
5. Do the substantial behavioral patterns of male and female students differ?

METHOD

Participants

The participants in this study were 52,110 students (29,273 Grade 9 students and 22,837 Grade 10 students) with an average age of approximately 15 years. They were enrolled through purposive sampling from approximately 320 junior and senior high schools in Taiwan. They included 27,062 boys and 25,048 girls. All of the students had taken basic science courses, which were designed according to the Taiwanese Grade 1–9 Curriculum Guidelines for Science and Technology, including courses covering water and soil conservation and temperature and heat. Because they had not previously participated in the CPS assessment used in this study, their teachers instructed them on the basic CPS concepts before they participated; the students also participated in an exercise unit called the Wellness Program.

CPS Assessments in Science Scenarios

Two science scenario units for CPS assessment were considered in this study (Kuo, 2014). The first was the Slurpee (**Figure 1**), which included a test taker and a computer agent. In this unit, the test taker collaborated with the computer agent to implement cryogenics using various constituents, such as salt, sugar, and monosodium glutamate, and ice. In addition, they were instructed to consider the proportion of cryogen to ice and attempt to determine the optimal proportions. The unit contained 17 multiple choice items, and each item corresponded to a target CPS skill (**Table 1**). The left panel of **Figure 1** illustrates the chatbox. The test taker was able to communicate with a simulated agent and collaborate to determine the optimal proportions of a cryogen. The left-down block included three alternatives, which were designed according to one of the CPS skills in **Table 1**. After the test taker selected an alternative, the choice was displayed in the history area of the chatbox. Responses from the computer agent, which were designed by experts, were displayed on the basis of the selected alternative. The right side of the screen (**Figure 1**) presented related materials such as figures, tables and notes. For example, the right screen in **Figure 1** shows three cryogen candidates and ice. In general, for each item, two dialog layers were used to create convergence or rescue points based on the responses of the computer agent. The first item of the Slurpee scenario is described as follows. The corresponding CPS skill was "(A1) Discovering perspectives and abilities of team members." Hence, the most appropriate solution was to ask the computer agent about the method of making ice.



Figure 1. Screenshot of the Slurpee unit (Kuo, 2014)
 Note: From Teachers' Collaborative Problem-Solving Teaching Competency Project, by B. C. Kuo, 2014

After the directions for the task, a message from computer agent reading “Let’s start to make Slurpee” was displayed. Then, three options were available for the test taker.

Item 1

Computer agent: Let’s start to make a Slurpee.

- I1M1 Let’s discuss the principles and methods of ice making.
- I1M2 I have never done it.
- I1M3 Everything is followed by you.

The credited response was I1M1. If the test taker selected I1M1, then they were able to discover the perspectives and abilities of the computer agent. Hence, they received 2 points for Item 1. Moreover, the response of the computer agent was “I1M1(R) I have found some information on the cooling, you can use the refrigerant method.” However, if the test taker did not select the credited response, the computer agent rescues and the test taker is presented with additional alternatives to consider. For example, a test taker selecting “I1M2 I have never done it” is insufficient evidence to determine that they lack CPS skill A1. Hence, the test takers were presented with additional options. The responses for I1M2(R) and I1M3(R) were identical: “I1M2(R) Computer agent: I would like to know our understanding of the ice making method.” Then, the following additional alternatives were presented.

- I1M21 I have no idea about ice making.
- I1M22 What do you know about the ice making method?
- I1M23 It will be difficult without a refrigerator.

If test taker selects I1M22, then they have demonstrated the A1 ability through the additional options. Hence, the test taker receives 1 point for Item 1. However, if they select I1M21, then they receive 0 points and the conversation of Item 1 is converged by the computer agent response, “I1M21(R) I have checked some information on the ice making. We can try the refrigerant method.” Because the task instructions stated that a refrigerator cannot be used, the test taker also receives 0 points and the response is the same. The additional alternatives may differ with respect to the options at the first stage. Thus, the following are the seven paths in Item 1.

1. Item 1 → I1M1 → I1M1(R) → Item 2;
2. Item 1 → I1M2 → I1M2(R) → I1M21 → I1M21(R) → Item 2;
3. Item 1 → I1M2 → I1M2(R) → I1M22 → I1M21(R) → Item 2;
4. Item 1 → I1M2 → I1M2(R) → I1M23 → I1M21(R) → Item 2;

5. Item 1 → I1M3 → I1M2(R) → I1M21 → I1M21(R) → Item 2;
6. Item 1 → I1M3 → I1M2(R) → I1M22 → I1M21(R) → Item 2;
7. Item 1 → I1M3 → I1M2(R) → I1M23 → I1M21(R) → Item 2.

This design was employed to ensure that the test taker can progress through the tasks with the computer agent despite selecting incorrect or nonoptimal alternatives. The student responses for this unit yielded a Cronbach's α of 0.82, which was between 0.80 and 0.90 and thus indicated excellent interrater reliability (De Vellis, 1991).

The Water Purification unit (Figure 2) involved a test taker, a highly collaboratively oriented computer agent and a less collaboratively oriented computer agent. In this 13-item unit, the test taker's role was to lead a group (two computer agents) to design an effective water purifier using the provided tools and thus win a competition. Cronbach's α was 0.74 for this unit, which was between 0.7 and 0.8 and thus indicated reasonable interrater reliability (De Vellis, 1991).

According to the PISA 2015 draft CPS framework (OECD, 2013) and the PISA 2015 Released Field Trial Cognitive Items (OECD, 2016), a single unit could not cover all 12 CPS skills because of the content of the corresponding science scenario. Hence, this study employed two units to test all the 12 CPS skills in science scenarios. However, if the items involved matrix component D3 ("Monitoring, providing feedback, and adapting the team organization and roles"), then the alternatives of the item were designed to encourage students to reflect on the process and to affirm the importance of attending to provided criteria to solve the problem efficiently (OECD, 2016). For example, test takers can provide reflective feedback on the work or suggests a collaborative method to improve their CPS performance (OECD, 2013). Because the CPS units contained multiple paths, three scores of 0, 1, and 2 were difficult to predetermine on the basis of chatting paths for D3. The corresponding scores depended on previous conversations (chatting paths) with the agents and the feedback options for the item. Because the scores should be determined according to the chatting paths through the CPS unit, the analysis of these items with respect to D3 was omitted in this study. Cronbach's α was 0.86 for the two units combined, which was between 0.80 and 0.90 and thus indicated excellent interrater reliability (De Vellis, 1991).

Coding Scheme

To observe behavioral changes over the testing period, the CPS target skills corresponding to the items were viewed as a CPS behavioral sequence. Moreover, a sequential analysis was applied to explore the significance

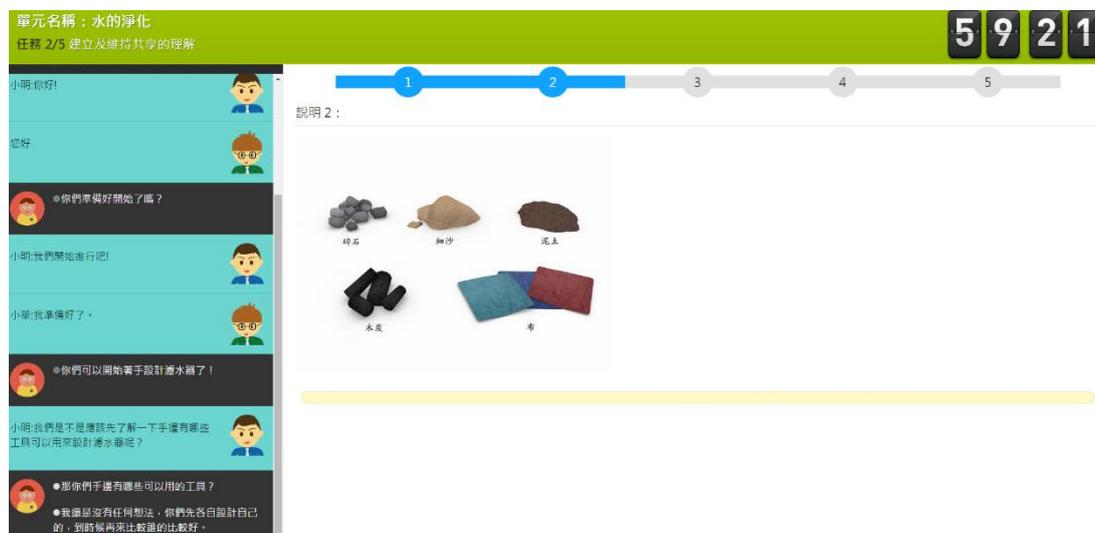


Figure 2. Screenshot of the Water Purification unit (Kuo, 2014)

Note. From Teachers' Collaborative Problem-Solving Teaching Competency Project, by B. C. Kuo, 2014

Table 2. Target CPS Skills of the Items

The Slurpee unit		Water Purification unit	
Item	Target CPS skill	Item	Target CPS skill
1	A1	1	A2
2	C1	2	A1
3	C2	3	B1
4	B2	4	C1
5	C1	5	C3
6	A3	6	D1
7	B3	7	C2
8	B1	8	D2
9	C2	9	D1
10	C3	10	C2
11	D2	11	D3
12	D1	12	D3
13	C1	13	D3
14	B1		
15	C2		
16	D2		
17	D3		

of the interactions among the CPS skills within the CPS test (Bakeman & Quera, 1995; Cheng & Hou, 2015). Because each item corresponded to a target CPS skill (Table 2), the order of the CPS behavioral sequence was fixed as follows:

A1C1C2B2C1A3B3B1C2C3D2D1C1B1C2D2 A2A1B1C1C3D1C2D2D1C2

A space was added between D2 and A2 because the former sequence is of the Slurpee unit and the latter is of the Water Purification unit. Hence, the behavioral pattern D2A2 cannot be included in the sequential analysis. Individual student CPS behavioral sequences could be distinguished by considering students' CPS skill proficiencies, which varied according to their response paths. The coding scheme depended on each student's CPS skill proficiency, as proposed in this section.

Suppose X and Y are two CPS skills in Table 1 that correspond to two adjacent items, i and j . For example, $X = B1$ and $Y = C2$ are respectively the CPS target skills of two adjacent items, namely items 8 and 9, in the Slurpee unit. In the two science scenarios in the CPS assessment, student CPS skill proficiency for X and Y were scored on a discrete scale, with possible values of 0, 1, and 2 (Kuo, 2014; Li et al., 2015). Hence, there are nine different paths based on received scores (Table 3). We divided these nine paths into three distinct categories (i.e., codes). If X and Y result in scores of 2, then the student has a high CPS skill proficiency with respect to the two adjacent items. In addition, this indicates that the student is highly efficient in CPS skill Y after effectively performing CPS skill X . Hence, the path of $X \rightarrow Y$ yields a score of 2. In other words, in sequential analysis, the frequency of XY should be higher than the frequency with respect to the other paths in Table 3. Therefore, we computed XY twice as shown in the third column in Table 3. A space was added between these two XY codes because they should be counted twice in the sequence. In the second category, the level of $X \rightarrow Y$ was 1, indicating that at least one instance of X and Y resulted in a score of 1 and that none of scores for X and Y were 0. In this case, the level of $X \rightarrow Y$ should be lower than that of the previous category, that is, the level of $X \rightarrow Y$ should be 2. Moreover, students can still obtain the score of the CPS skill even if the score is lower than that of the previous category. Hence, the level of $X \rightarrow Y$ was set to 1. Finally, if one of the skill proficiency values of X and Y was 0, then no convincing evidence indicated that the student can smoothly progress from X to Y . Therefore, the level of $X \rightarrow Y$ was scored as 0.

Table 3. Coding Scheme for the Nine Paths from X to Y Based on Scores

X	Y	The level of X → Y	Code
2	2	2	XY XY
2	1	1	XY
2	0	0	N/A
1	2	1	XY
1	1	1	XY
1	0	0	N/A
0	2	0	N/A
0	1	0	N/A
0	0	0	N/A

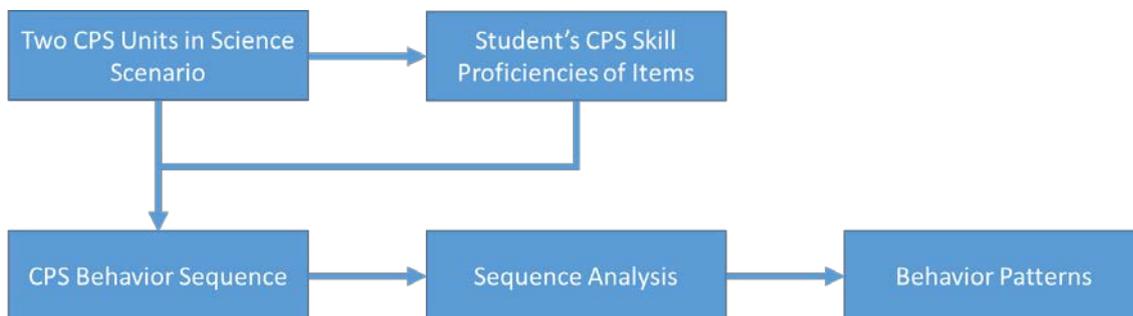


Figure 3. Research flowchart

Sequence analysis can indicate the level of dependence (significance level) between X and Y (Bakeman & Quera, 1995). If X and Y are dependent, then $P(XY) \neq P(X)P(Y)$. The transition frequencies of XY were calculated, and a variation between observed and expected frequencies (z score) of >1.96 indicated that the given sequential pattern was statistically significant (i.e., $X \rightarrow Y$) (Cheng & Hou, 2015). In this study, the p value of the z score was also computed. Y featured a connection from X (i.e., $X \rightarrow Y$) if the corresponding p value reached a significance level of .001.

Figure 3 exhibits the research flowchart for this study. First, student CPS skill proficiencies and target CPS skills for the items were employed to define the CPS behavioral sequence. Then, sequence analysis was then implemented to examine the significant behavioral patterns of students.

Table 4 provides an example student response for the Slurpee unit. Items i and j indicate two adjacent items corresponding to the target CPS skills X and Y. The fifth column in **Table 4** lists the proposed coding scheme. Spaces were added in the sequences to ensure that the behavioral pattern was counted twice. Therefore, the corresponding coded behavioral sequences for this student and unit were as follows:

A1C1C2 C1C2B2 C2B2C1 B2C1A3B3 C3D2C3D2 C2D2

Table 4. Example Student Response and Corresponding Codes for Two Adjacent Items

Item <i>i</i>	Item <i>j</i>	<i>X</i>	<i>Y</i>	Code
1	2	A1 1	C1 2	A1C1
2	3	C1 2	C2 2	C1C2 C1C2
3	4	C2 2	B2 2	C2B2 C2B2
4	5	B2 2	C1 2	B2C1 B2C1
5	6	C1 2	A3 1	C1A3
6	7	A3 1	B3 2	A3B3
7	8	B3 2	B1 0	N/A
8	9	B1 0	C2 0	N/A
9	10	C2 0	C3 2	N/A
10	11	C3 2	D2 2	C3D2 C3D2
11	12	D2 2	D1 0	N/A
12	13	D1 0	C1 2	N/A
13	14	C1 2	B1 0	N/A
14	15	B1 0	C2 1	N/A
15	16	C2 1	D2 1	C2D2

RESULTS AND DISCUSSION

CPS Skill Proficiency Analysis

In this section, we compare the student proficiency measures for the 11 CPS skills excluding D3; moreover, we compare collaborative competency and problem-solving competency. [Figure 4](#) presents the average proficiency scores for the Taiwanese students for the 11 CPS skills in the science scenarios. The average scores for the 11 CPS skills were all higher than 1. Each item was scored 0, 1, or 2, corresponding to three respective levels for the target CPS skills. Hence, on average, the Taiwanese students were at least proficient in these 11 skills. The two highest average scores were received for B3 and C1, indicating that the Taiwanese students were efficient in describing the roles of team members (B3) and communicating with team members (C1) about actions. However, the skill of building a shared representation or negotiating the meaning of a problem with other team members (B1) was lacking in Taiwanese students.

Furthermore, the collaborative competencies, (1), (2), and (3), and the problem-solving competencies, (A), (B), (C), and (D), could be computed by averaging the CPS scores for related items ([Figure 5](#)). All of them had scores higher than 1. The highest score (1.69) was obtained for problem-solving competency (A). Thus, students could effectively discuss with team members how to solve the given problem (OECD, 2013). Although the score for problem-solving competency (B) was the lowest, it was similar to the scores for competencies (C), (D), and (2). [Table 5](#) lists the correlations between the collaborative and problem-solving competencies, most of which had

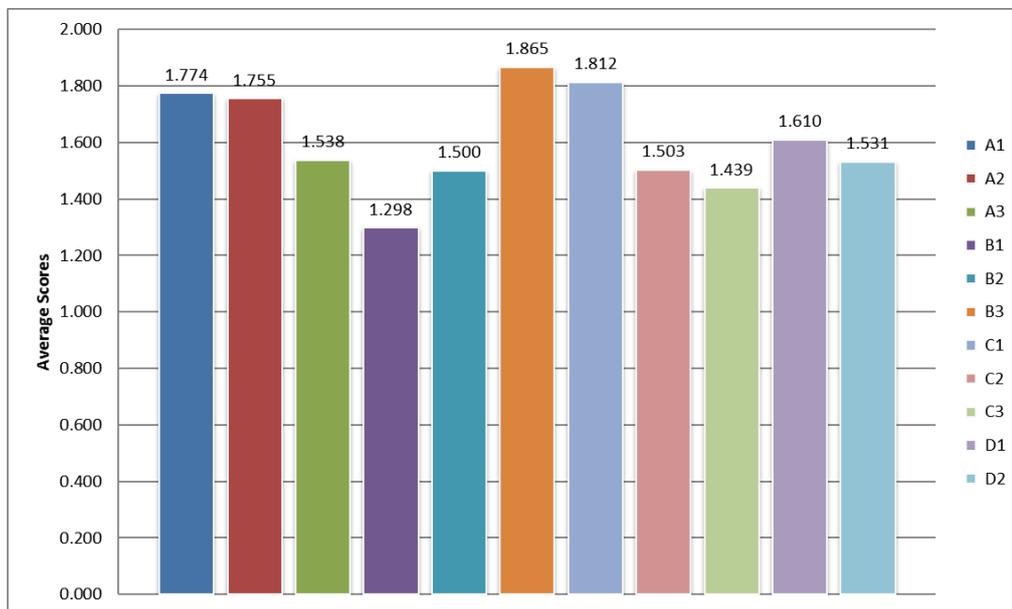


Figure 4. Average proficiency scores for the Taiwanese students for the 11 CPS skills in science scenarios

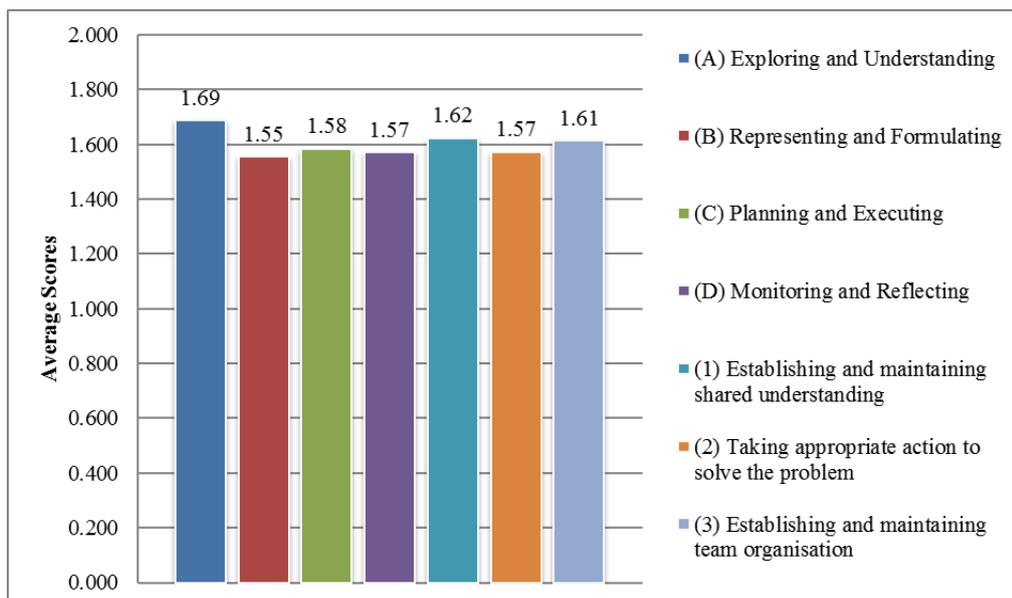


Figure 5. Average scores for collaborative competencies, (1), (2), and (3), and problem-solving competencies, (A), (B), (C), and (D)

scores higher than 0.7; therefore, most collaborative and problem-solving competencies were highly correlated, except for (B) and (3), which were moderately correlated.

The independent sample test results for the male and female students' 11 CPS skill proficiencies are provided in Table 6. The variances between the male and female students for the 11 CPS skill proficiencies were unequal because the significance levels corresponding to the test for equality of variance were less than .001. Hence, all *t* test statistics in Table 6 assume unequal variance. Moreover, because the *p* values were less than .001, female student CPS skill proficiencies were significantly higher than those of male students.

Table 5. Correlation between Collaborative Competencies and Problem-Solving Competencies

Correlation	(1) Establishing and maintaining shared understanding	(2) Taking appropriate action to solve the problem	(3) Establishing and maintaining team organisation
(A) Exploring and Understanding	0.724	0.734	0.763
(B) Representing and Formulating	0.711	0.709	0.612
(C) Planning and Executing	0.734	0.714	0.797
(D) Monitoring and Reflecting	0.781	0.724	N/A

Table 6. Independent Sample Test Results for the 11 CPS Skill Proficiencies of Male and Female Students

CPS Skill	Sex	Average Score	Test for Equality of Variances			t-test for Equality of Means	
			F	p-value	Equality	t	p-value
A1	M	1.73	778.90	0.00	No	-28.170	0.00
	F	1.82					
A2	M	1.69	750.784	0.00	No	-28.17	0.00
	F	1.83					
A3	M	1.45	750.79	0.00	No	-34.80	0.00
	F	1.63					
B1	M	1.27	165.02	0.00	No	-12.88	0.00
	F	1.33					
B2	M	1.46	299.7	0.00	No	-17.36	0.00
	F	1.55					
B3	M	1.81	728.55	0.00	No	-27.41	0.00
	F	1.92					
C1	M	1.76	1147.22	0.00	No	-34.33	0.00
	F	1.86					
C2	M	1.45	1238.29	0.00	No	-35.36	0.00
	F	1.56					
C3	M	1.40	313.75	0.00	No	-17.78	0.00
	F	1.48					
D1	M	1.57	593.63	0.00	No	-24.5	0.00
	F	1.65					
D2	M	1.47	1099.39	0.00	No	-33.4	0.00
	F	1.60					

Note. *p < .05, **p < .01, ***p < .001

Sequence Analysis

For an improved understanding of the transit between the target CPS skills for two adjacent items, a series of sequence analyses were applied to explore the behavioral patterns for (a) all students, (b) all male students, (c) all female students, (d) all students in a given class, (e) all male students in a given class, and (f) all female students in a given class. Because the sequence of target CPS skills corresponding to items was fixed, only 20 behavioral patterns were observed (Figure 6) for the two science scenario units.

Table 7 lists the adjusted residuals for the sequence analysis of all students. Each z score (p value) indicates whether a path between two related CPS skills was dependent. The row categories represent the starting behavior, and the column categories indicate the subsequent behavior. For example, in the second row and first column of Table 7, A2 is the starting behavior and A1 is the subsequent behavior. The corresponding z value is 1183.933, and the p value is .000, which is less than .001. Thus, the sequence A2 → A1 was significant. Hence, in the corresponding behavioral transition diagram, a black arrow points from A2 to A1. "N/A" in Table 7 indicates that the corresponding CPS skill behavioral pattern was not detected in the two units. In the CPS skill sequences combining

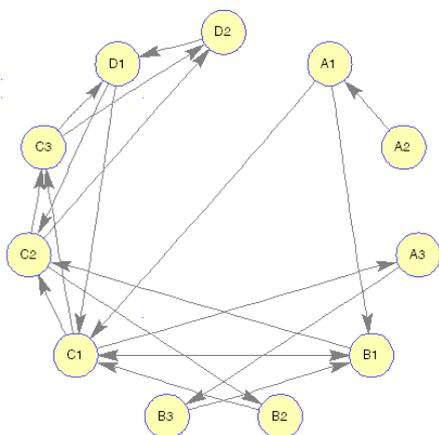


Figure 6. The 20 observed behavioral patterns in the two science scenario units

Table 7. Adjusted Residuals Table for Sequence Analysis of All Students (*p* value shown in brackets)

z (p-value)	A1	A2	A3	B1	B2	B3	C1	C2	C3	D1	D2
A1	N/A	N/A	N/A	416.499 (0.000)	N/A	N/A	456.458 (0.000)	N/A	N/A	N/A	N/A
A2	1183.933 (0.000)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
A3	N/A	N/A	N/A	N/A	N/A	1337.748 (0.000)	N/A	N/A	N/A	N/A	N/A
B1	N/A	N/A	N/A	N/A	N/A	N/A	193.429 (0.000)	378.544 (0.000)	N/A	N/A	N/A
B2	N/A	N/A	N/A	N/A	N/A	N/A	602.615 (0.000)	N/A	N/A	N/A	N/A
B3	N/A	N/A	N/A	751.493 (0.000)	N/A	N/A	N/A	N/A	N/A	N/A	N/A
C1	N/A	N/A	582.536 (0.000)	220.608 (0.000)	N/A	N/A	N/A	138.620 (0.000)	248.619 (0.000)	N/A	N/A
C2	N/A	N/A	N/A	N/A	519.449 (0.000)	N/A	N/A	N/A	224.773 (0.000)	N/A	528.117 (0.000)
C3	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	326.232 (0.000)	599.859 (0.000)
D1	N/A	N/A	N/A	N/A	N/A	N/A	167.994 (0.000)	644.189 (0.000)	N/A	N/A	N/A
D2	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	805.409 (0.000)	N/A

Note. **p* < .05, ***p* < .01, ****p* < .001

all students' behavioral sequences, all 20 paths were connected and significant. This was because, on average, the Taiwanese students were at least proficient in the 11 CPS skills (Figure 4). Thus, the CPS behavioral transition diagram for all students corresponded to the results provided in Figure 6. To examine the differences in the CPS behavioral patterns of the male and female students, the adjusted residuals tables for their sequence analyses are shown in Tables 8 and 9, respectively. The results were similar to those for all the students. All 20 behavioral patterns were significant.

Table 8. Adjusted Residuals Table for Sequence Analysis of Male Students (*p* value shown in brackets)

z (p-value)	A1	A2	A3	B1	B2	B3	C1	C2	C3	D1	D2
A1	N/A	N/A	N/A	284.646 (0.000)	N/A	N/A	324.462 (0.000)	N/A	N/A	N/A	N/A
A2	825.367 (0.000)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
A3	N/A	N/A	N/A	N/A	N/A	932.164 (0.000)	N/A	N/A	N/A	N/A	N/A
B1	N/A	N/A	N/A	N/A	N/A	N/A	132.837 (0.000)	250.753 (0.000)	N/A	N/A	N/A
B2	N/A	N/A	N/A	N/A	N/A	N/A	424.741 (0.000)	N/A	N/A	N/A	N/A
B3	N/A	N/A	N/A	543.478 (0.000)	N/A						
C1	N/A	N/A	407.632 (0.000)	158.176 (0.000)	N/A	N/A	N/A	97.744 (0.000)	182.964 (0.000)	N/A	N/A
C2	N/A	N/A	N/A	N/A	364.779 (0.000)	N/A	N/A	N/A	140.248 (0.000)	N/A	373.362 (0.000)
C3	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	236.433 (0.000)	417.932 (0.000)
D1	N/A	N/A	N/A	N/A	N/A	N/A	110.797 (0.000)	468.500 (0.000)	N/A	N/A	N/A
D2	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	559.549 (0.000)	N/A

Note. **p* < .05, ***p* < .01, ****p* < .001

Table 9. Adjusted Residuals Table for Sequence Analysis of Female Students (*p* value shown in brackets)

z (p-value)	A1	A2	A3	B1	B2	B3	C1	C2	C3	D1	D2
A1	N/A	N/A	N/A	304.290 (0.000)	N/A	N/A	321.097 (0.000)	N/A	N/A	N/A	N/A
A2	848.834 (0.000)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
A3	N/A	N/A	N/A	N/A	N/A	959.764 (0.000)	N/A	N/A	N/A	N/A	N/A
B1	N/A	N/A	N/A	N/A	N/A	N/A	140.675 (0.000)	284.417 (0.000)	N/A	N/A	N/A
B2	N/A	N/A	N/A	N/A	N/A	N/A	427.482 (0.000)	N/A	N/A	N/A	N/A
B3	N/A	N/A	N/A	519.333 (0.000)	N/A						
C1	N/A	N/A	416.269 (0.000)	153.831 (0.000)	N/A	N/A	N/A	98.295 (0.000)	168.790 (0.000)	N/A	N/A
C2	N/A	N/A	N/A	N/A	369.815 (0.000)	N/A	N/A	N/A	177.270 (0.000)	N/A	373.564 (0.000)
C3	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	225.006 (0.000)	430.291 (0.000)
D1	N/A	N/A	N/A	N/A	N/A	N/A	126.740 (0.000)	442.590 (0.000)	N/A	N/A	N/A
D2	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	579.433 (0.000)	N/A

Note. **p* < .05, ***p* < .01, ****p* < .001

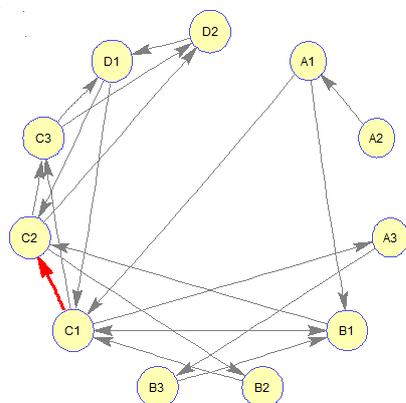


Figure 7. CPS behavioral transition diagram for 22 students in a class from northern Taiwan

Table 10. Adjusted Residuals Table for Sequence Analysis of 22 Students in a Class from Northern Taiwan (*p* value shown in brackets)

<i>z</i> (<i>p</i> -value)	A1	A2	A3	B1	B2	B3	C1	C2	C3	D1	D2
A1	N/A	N/A	N/A	8.023 (0.000)	N/A	N/A	10.018 (0.000)	N/A	N/A	N/A	N/A
A2	26.814 (0.000)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
A3	N/A	N/A	N/A	N/A	N/A	27.503 (0.000)	N/A	N/A	N/A	N/A	N/A
B1	N/A	N/A	N/A	N/A	N/A	N/A	3.144 (0.000)	9.390 (0.000)	N/A	N/A	N/A
B2	N/A	N/A	N/A	N/A	N/A	N/A	13.391 (0.000)	N/A	N/A	N/A	N/A
B3	N/A	N/A	N/A	17.373 (0.000)	N/A	N/A	N/A	N/A	N/A	N/A	N/A
C1	N/A	N/A	13.421 (0.000)	4.880 (0.000)	N/A	N/A	N/A	2.578 (0.005)	6.803 (0.000)	N/A	N/A
C2	N/A	N/A	N/A	N/A	11.282 (0.000)	N/A	N/A	N/A	4.417 (0.000)	N/A	12.665 (0.000)
C3	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	8.916 (0.000)	11.277 (0.000)
D1	N/A	N/A	N/A	N/A	N/A	N/A	4.425 (0.000)	12.246 (0.000)	N/A	N/A	N/A
D2	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	16.070 (0.000)	N/A

Note. **p* < .05, ***p* < .01, ****p* < .001

Sequence analysis can be applied not only to all participating students but also to the students in a given class. This type of report can provide their teachers with constructive feedback on their learning activities and enable them to train students that lack sufficient CPS behavioral transition patterns in science courses. **Table 10** shows the adjusted residuals table for the sequence analysis of 22 students (13 female and 9 male students) in a class from northern Taiwan; the corresponding CPS behavioral transition diagram for the class is presented in **Figure 7**. In the seventh row and eighth column of **Table 10** (gray cell), the *p* value is .005. Thus, the sequence C1 → C2 was nonsignificant in this study. Hence, a red arrow pointing from C1 to C2 is used to highlight the behavioral pattern in **Figure 7**. Overall, students in this class were not proficient in enacting plans (C2) after communicating with team members about the actions (C1). Therefore, the science teacher can design further related collaborative activities in science class to train students.

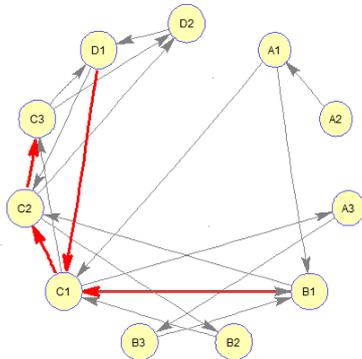


Figure 8. CPS behavioral transition diagram for male students in the class

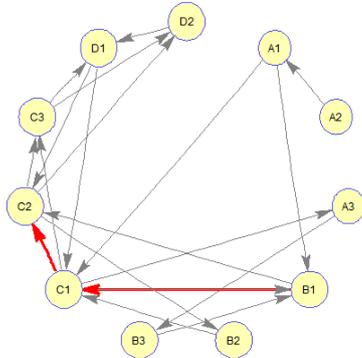


Figure 9. CPS behavioral transition diagram for female students in the class

The CPS behavioral patterns of the male and female students in the class are illustrated in **Figures 8** and **9**, respectively. **Tables 11** and **12** respectively comprise the corresponding adjusted residuals tables for sequence analysis. **Figure 8** and **Table 11** demonstrate that the male students in the class were not proficient in behavioral patterns $C1 \rightarrow C2$, $C2 \rightarrow C3$, $D1 \rightarrow C1$, and $B1 \rightarrow C1$. Moreover, **Figure 9** and **Table 12** indicate that the female students in the class were not proficient in behavioral patterns $C1 \rightarrow C2$ and $B1 \rightarrow C1$. Thus, the male students lacked more behavioral patterns than the female students in the class did. Additionally, comparing **Figures 8** and **9** with **Figure 7** indicates that some students, regardless of gender, were not proficient in communicating with team members about actions after building a shared representation and negotiating the meaning of the problem.

CONCLUSION

In this study, we analyzed the CPS skill proficiencies of Taiwanese students (aged approximately 15 years) in science scenarios, and we explored their behavioral patterns through sequential analysis. The results demonstrated that the Taiwanese students were at least proficient in the 11 CPS skills. They exhibited particularly strong proficiency in B3 and C1 and low proficiency in B1. Studies have demonstrated that, on average, male students exhibit higher learning achievement than female students in science and that the gender differences have decreased (Browne & Ross, 1991; Murphy, 1997; Ziegler & Heller, 1997; Kimbell, Stables, Wheeler, Wosniak, & Kelly, 1991). However, female students tend to work collaboratively with others, whereas male students tend to work independently (Kimbell et al., 1991). **Table 6** indicates that the CPS skill proficiencies for the female students were significantly superior to those of the male students for all 11 CPS skills. The finding is in agreement with those of Kimbel et al. (1991).

Table 11. Adjusted Residuals Table for Sequence Analysis of Male Students in the Class (*p* value shown in brackets)

z (p-value)	A1	A2	A3	B1	B2	B3	C1	C2	C3	D1	D2
A1	N/A	N/A	N/A	5.197 (0.000)	N/A	N/A	6.311 (0.000)	N/A	N/A	N/A	N/A
A2	16.889 (0.000)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
A3	N/A	N/A	N/A	N/A	N/A	16.947 (0.000)	N/A	N/A	N/A	N/A	N/A
B1	N/A	N/A	N/A	N/A	N/A	N/A	1.506 (0.066)	5.961 (0.000)	N/A	N/A	N/A
B2	N/A	N/A	N/A	N/A	N/A	N/A	8.859 (0.000)	N/A	N/A	N/A	N/A
B3	N/A	N/A	N/A	11.326 (0.000)	N/A	N/A	N/A	N/A	N/A	N/A	N/A
C1	N/A	N/A	8.631 (0.000)	3.656 (0.000)	N/A	N/A	N/A	1.778 (0.038)	5.112 (0.000)	N/A	N/A
C2	N/A	N/A	N/A	N/A	7.048 (0.000)	N/A	N/A	N/A	2.191 (0.014)	N/A	7.809 (0.000)
C3	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	6.621 (0.000)	7.562 (0.000)
D1	N/A	N/A	N/A	N/A	N/A	N/A	1.969 (0.024)	8.141 (0.000)	N/A	N/A	N/A
D2	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	9.702 (0.000)	N/A

Note. **p* < .05, ***p* < .01, ****p* < .001

Table 12. Adjusted Residuals Table for Sequence Analysis of Female Students in the Class (*p* value shown in brackets)

z (p-value)	A1	A2	A3	B1	B2	B3	C1	C2	C3	D1	D2
A1	N/A	N/A	N/A	6.162 (0.000)	N/A	N/A	7.782 (0.000)	N/A	N/A	N/A	N/A
A2	20.772 (0.000)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
A3	N/A	N/A	N/A	N/A	N/A	21.747 (0.000)	N/A	N/A	N/A	N/A	N/A
B1	N/A	N/A	N/A	N/A	N/A	N/A	2.882 (0.002)	7.291 (0.000)	N/A	N/A	N/A
B2	N/A	N/A	N/A	N/A	N/A	N/A	10.066 (0.000)	N/A	N/A	N/A	N/A
B3	N/A	N/A	N/A	13.143 (0.000)	N/A	N/A	N/A	N/A	N/A	N/A	N/A
C1	N/A	N/A	10.278 (0.000)	3.273 (0.000)	N/A	N/A	N/A	1.877 (0.030)	4.520 (0.000)	N/A	N/A
C2	N/A	N/A	N/A	N/A	8.796 (0.000)	N/A	N/A	N/A	4.054 (0.000)	N/A	9.954 (0.000)
C3	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	6.075 (0.000)	8.458 (0.000)
D1	N/A	N/A	N/A	N/A	N/A	N/A	4.095 (0.000)	9.165 (0.000)	N/A	N/A	N/A
D2	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	12.785 (0.000)	N/A

Note. **p* < .05, ***p* < .01, ****p* < .001

Overall, regardless of gender, among the Taiwanese students, all 20 of the detected sequential patterns in science scenarios were significant. However, if a single class is considered, some behavioral patterns may be lacking. In the case of a class from northern Taiwan, most of the students were not proficient in enacting plans after communicating with team members. Moreover, some students in the class could not effectively communicate with team members regarding the actions (Figure 7). In addition, we compared the significant behavioral patterns of male and female students in the class and revealed that the male students exhibited low performance in C2 → C3 and D1 → C1. We hope that the findings regarding CPS skill proficiencies and CPS behavioral transition patterns in this study can serve as a reference for teachers, researchers, and governmental departments for designing future training courses or government policies.

ACKNOWLEDGEMENTS

Authors would like to thank the participants of this study. We are also grateful to the team of Dr. Bor-Chen Kuo for providing the CPS assessment system supported by the Ministry of Education, Taiwan. Moreover, we acknowledge Wallace Academic Editing for editing this manuscript. This work was supported by the Ministry of Science and Technology, Taiwan, under Grant Most 103-2511-S-142-012-MY3 and 106-2511-S-142-003-MY3.

REFERENCES

- Baines, E., Blatchford, P., & Kutnick, P. (2003). Changes in grouping practice over primary and secondary school. *International Journal of Educational Research*, 39, 9-34.
- Bakeman, R., & Quera, V. (1995). *Analyzing Interaction; Sequential analysis with SDIS and GSEQ*. New York, NY: Cambridge University Press.
- Brannick, M. T., & Prince, C. (1997) An overview of team performance measurement. In M. T. Brannick, E. Salas, & C. Prince (Eds.), *Team performance assessment and measurement: Theory methods and applications* (pp. 3-16). Mahwah, NJ; Lawrence Erlbaum Associates.
- Browne, N. & Ross, C. (1991). Girls' stuff, boys' stuff: Young children talking and playing. In N. Browne (Ed.), *Science and technology in the early years*. Buckingham: Open University Press.
- Cheng, K. H., & Hou, H. T. (2015). Exploring students' behavioural patterns during online peer assessment form the affective, cognitive, and metacognitive perspectives: a progressive sequential analysis, *Technology, Pedagogy, and Education*, 24(2), 171-188.
- Coleman, E. (1998). Using explanatory knowledge during collaborative problem solving in science. *Journal of Learning Sciences*, 7, 387-427
- De Vellis, R.F. (1991). *Scale Development: theory and applications*. Newbury Park: Sage.
- Duschl, R., & Osborne, J. (2002). Supporting and promoting argumentation discourse. *Studies in Science Education*, 38, 39-72.
- Fiore, S. M., Rosen, M., Salas, E., Burke, S., & Jentsch, F. (2008). Processes in complex team problem solving: Parsing and defining the theoretical problem space. In M. Letsky, N. Warner, S. M. Fiore, & C. Smith (Eds.). *Macrocognition in Teams: Theories and Methodologies*. London: Ashgate Publishers.
- Fiore, S., Rosen, M., Smith-Jentsch, K., Salas, E., Letsky, M. & Warner, N. (2010). Toward an understanding of macrocognition in teams: Predicting process in complex collaborative contexts. *The Journal of the Human Factors and Ergonomics Society*, 53, 203-224.
- Franklin, S. & A.C. Graesser (1996). "Is it an agent or just a program? A taxonomy for autonomous agents," *Proceedings of the Agent Theories, Architectures, and Languages Workshop*, Springer-Verlag, Berlin.
- Funke, J. & Frensch, P. A. (2007). Complex problem solving: The European perspective - 10 years after. In D. H. Jonassen (Ed.), *Learning to Solve Complex Scientific Problems* (pp. 25-47). New York: Lawrence Erlbaum.
- Griffin, P., Care, E., & McGaw (2011). The changing role of education and schools. In P. Griffin, B. McGaw, & E. Care (Eds.). *Assessment and teaching 21st century skills* (pp. 1-15). Heidelberg: Springer.
- Griffin, P., McGaw, B. & Care, E. (2011). *Assessment and teaching 21st century skills*. Heidelberg: Springer.

- Healy, A. F., Buck-Gengler, C. J., Barshi, I., Parker, J. T., Schneider, V. L., Raymond, W. D., LaVoie, N. N., Bowles, A. R., Pauli, P., Fisher, J. A., & Bourne, L. E., Jr. (2002). Optimizing the durability and generalizability of knowledge and skills. In S. P. Shohov (Ed.), *Advances in psychology research* (Vol. 8, pp. 103-174). Huntington, NY: Nova Science Publishers
- Hou, H. T. (2010). Explore the behavioural patterns in project-based learning with online discussion: Quantitative content analysis and progressive sequential analysis. *Trkish Online Journal of Educational Technology*, 9(3), 52-60.
- Hou, H. T., Chang, K. E., & Sung, Y. T. (2007a). An Analysis of Peer Assessment Online Discussions with Course that uses Project-Based Learning. *Interactive Learning Environments*, 15(3), 237-251.
- Hou, H. T., Chang, K. E., & Sung, Y. T. (2007b, July). *Analysis of Time-Management Pattern of Interactive Behaviors during Online Project-Based Learning*, Paper presented at International Conference on Advanced Learning Technologies, Nigatta, Japan.
- Hou, H. T., Chang, K. E., & Sung, Y. T. (2008). Analysis of Problem-Solving Based Online Asynchronous Discussion Pattern. *Educational Technology & Society*, 11(1), 17-28.
- Hou, H. T., Chang, K. E., & Sung, Y. T. (2009). Using Bolgs as a Professional Development Tool for Teachers: Analysis of Interaction Behavioral Patterns, *Interactive Learning Environments*, 17(4), 325-340.
- Kimbell, R., Stables, K., Wheeler, T., Wosniak, A., & Kelly, V. (1991). *The assessment of performance in design and technology: Final report*. London: Schools Examination and Assessment Council.
- Kuo, B.-C. (2014). Teachers' collaborative problem solving teaching competency project (in Chinese). Retrieved from <https://sites.google.com/site/cpswebsite2014/>
- Kutnick, P. & Blatchford, P. (2013). *Effective group-work in Primary school classrooms: the SPRinG approach*. Dordrecht: Springer.
- Lee, A.Y., & Pennington, N. (1993). *The effect of experience on a cross-domain transfer of diagnostic skill*. Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society. Hillsdale, NJ: Erlbaum.
- Li, C. H., Pai, K. C., Kuo, B. C., Lin, Y.N., & Liu, Z. Y. (2015). *The development of online Chinese collaborative problem solving tests in science*. Proceedings of 2015 International Meeting of the Psychometric Society. Beijing, China.
- Luckin, R., Baines, E., Cukurova, M., & Holmes, W. (2017). *Solved! Making the case for collaborative problem-solving*. A report for Nesta. Nesta, London, UK.
- Ma, X. (2008). Within-School Gender Gaps in Reading, Mathematics, and Science Literacy. *Comparative Education Review*, 52(3), 437-460.
- Mayer, R. E. & Wittrock, M. C. (1996). Problem solving transfer. In R. Calfee & R. Berliner (Eds.), *Handbook of Educational Psychology* (pp. 47-62). New York, NY: Macmillan.
- Mayer, R. E. (1992). *Thinking, Problem solving, Cognition* (2nd ed.). New York, NY: Freeman.
- Millis, K., Forsyth, C., Butler, H., Wallace, P., Graesser, A., & Halpern, D. (2011). Operation ARIES!: A Serious Game for Teaching Scientific Inquiry. In Ma, M., Oikonomou, A., & Jain, L. C. (Eds.), *Serious games and edutainment applications*, Springer-Verlag, London, UK.
- Murphy, P. (1997). Gender differences: Messages for science learning. In K. Harnquist and A. Bergen (Eds.), *Growing up with science: Developing early understanding of science*. London: Jessica Kingsley.
- National Research Council (2011). *Assessing 21st century skills*. Washington, DC: National Academies Press.
- O'Neil, H. F., Chuang, S. H., Baker, E. L. (2010). Computer-based feedback for computer-based collaborative problem-solving. In D. Ifenthaler, P. Pirnay-Dummer, N. M. Seel (Eds.), *Computer-based diagnostics and systematic analysis of knowledge*. New York, NY: Springer-Verlag.
- O'Neil, H. F., Chuang, S., & Chung, G. K. W. K. (2003). Issues in the computer- based assessment of collaborative problem solving. *Assessment in Education*, 10, 361-373.
- OECD (2013). *PISA 2015 Draft collaborative problem solving framework*. Retrived from <http://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Collaborative%20Problem%20Solving%20Framework%20.pdf>

- OECD (2016). *PISA 2015 Released Field Trial Cognitive Items*. Retrived from <http://www.oecd.org/pisa/pisaproducts/PISA2015-Released-FT-Cognitive-Items.pdf>
- OECD. (2003). *The PISA 2003 Assessment Framework: Mathematics, Reading, Science and Problem Solving Knowledge and Skills*. Retrived from <http://www.oecd.org/edu/preschoolandschool/programmeforinternationalstudentassessmentpisa/33694881.pdf>
- OECD. (2010). *PISA 2012 Field Trial Problem Solving Framework*. Retrived from <http://www.oecd.org/dataoecd/8/42/46962005.pdf>
- Rosen, Y., & Rimor, R. (2012). Teaching and assessing problem solving in online collaborative environment. In R. Hartshorne, T. Heafner, & T. Petty (Eds.), *Teacher education programs and online learning tools: Innovations in teacher preparation* (pp. 82-97). Hershey, PA: Information Science Reference.
- Salas, E., Dickenson, T. L., Converse, S. and Tannenbaum, S. I. (1992). Toward an understanding of team performance and training. In R. W. Swezey and E. Salas (Eds), *Teams: Their Training and Performance* (pp. 3-29). Norwood, NJ: Ablex.
- Salas, E., Cooke, N.J., & Rosen, M. A. (2008). On teams, teamwork, and team performance: discoveries and developments. *Human Factors*, 50, 540-548.
- Schroeder, C.M., Scott, T.P., Tolson, H., Huang, T. & Lee, Y. (2007). A meta-analysis of national research: Effects of teaching strategies on student achievement in science in the United States. *Journal of Research in Science Teaching*, 44(10), 1436-1460.
- Sjøberg, S., & Schreiner, C. (2005). How do learners in different cultures relate to science and technology? Results and perspectives from the project ROSE. *Asia-Pacific Forum on Science Learning and Teaching*, 6(2), 1-17.
- Webb, N. & Palincsar, A. (1996). Group processes in the classroom. In Berliner, D. C. & Calfee, R.C. (Eds.), *Handbook of educational psychology*. New York: Macmillan.
- Wiley, J., & Jensen, M.S. (2007). When small problem solving groups are effective: What leads to successful interactions? Symposium conducted at the Annual Meeting of the Society for Text & Discourse, Glasgow, Scotland.
- Yip, D. Y., Chiu, M. M. & Ho, E. S. C. (2004). Hong Kong student achievement in OECD-PISA study: Gender differences in science content, literacy skills, and test item formats. *International Journal of Science and Mathematics Education*, 2, 91-106.
- Ziegler, A. & Heller, K.A. (1997). Attribution retraining for self-related cognitions among women. *Gifted and Talented International*, 12(1), 36-41.

<http://www.ejmste.com>