Abstract

Collaborative problem solving (ColPS) skills are considered crucial to succeed in work, education, and life in a knowledge-rich society. Despite this relevance and the ample need for valid and reliable assessments to measure ColPS, research on the assessment of ColPS is at its initial stage. The present study attempts to fill this gap by developing and evaluating a novel ColPS task with the help of think-aloud protocols. The task was developed on the basis of a ColPS framework, and principles emphasized in the research literature on students' interaction, collaboration, and problem solving were implemented. A real-world problem mimicking a common teaching and learning situation formed the context of this task. The empirical evidence obtained from the think-aloud protocols of twelve Norwegian students displayed the strengths and weaknesses of the task, and strengthened the feasibility to assess ColPS. Implications for the future design of ColPS tasks are discussed.

Keywords: Assessment and teaching of 21st century skills (ATC21S); Collaborative problem solving; Computer-based assessment; Educational technology; Think-aloud protocol

This is a pre-print version.
DOI to published version: 10.1016/j.chb.2017.08.007

Revealing the Processes of Students' Interaction with a New Collaborative Problem Solving Task: An in-depth analysis of Think-Aloud Protocols

Introduction

Twenty-first century skills are increasingly regarded as critical in our complex and information-rich society (Griffin, Care & McGaw, 2012; P21, 2012). Several agent stakeholders emphasize the integration and development of skills that are related to problem solving, communication and collaboration, and the use of digital technology in the context of education (Law, Lee & Yuen, 2009; NRC, 2012; Quellmalz, 2009; Suto, 2013). As a consequence, assessments of these critical skills have gained considerable attention, and challenges related to their measurement have been pointed out (Griffin, McGaw, & Care, 2012; Griffin & Care, 2015). These challenges have particularly emerged in the measurement of collaborative problem solving (ColPS) – a skill that goes beyond students' individual problem solving capabilities as they "join their understandings and efforts and work together on solving […] problem situations" (OECD, 2013, p. 1). Silva (2009) described the current status of computer-based ColPS assessments as perplexing, and scholars have accentuated the limited empirical research on students' ColPS (Ras, Krkovic, Greiff, Tobias, & Maquil, 2014). This may be partly due to the complexities collaborative activities add to the methodological approaches that are used to describe individual students' skills (von Davier & Halpin, 2013). Within the 21st century skills research, it is particularly surprising that a very limited number of studies investigated the processes of direct student-student collaboration to uncover the complex processes involved in problem solving (Siddiq, Hatlevik, Olsen, Throndsen, & Scherer, 2016).

Researchers have argued that the ColPS construct is multifaceted and consist of two broader skillsets: social and cognitive skills (Care, Scoular, & Griffin, 2016). Even though there is an emergence of interest in assessing these approaches simultaneously, the capacity to capture the cognitive skillset has been more successful compared to the social skillset (Care et

This is a pre-print version.
DOI to published version: 10.1016/j.chb.2017.08.007

al., 2016). Yet, the challenge lies in “how to capture more of this (social skills) in an automated way that lends itself to coding and scoring” (Care et al., 2016, p. 24).

Against this background, the present study aims to provide insights into students’ interactions, communication, collaboration, and problem solving skills in a digital environment at lower-secondary level. The overarching aim is to investigate to what extent a newly developed ColPS task facilitates and captures students’ abilities to interact with each other and solve a problem within an assessment; moreover, we describe these processes on the basis of an underlying framework. We examine video data of twelve think-aloud protocols (TAP), while students were taking a performance-based test.

Theoretical Framework

Twenty-First Century Skills

The ubiquitous increase of ICT has affected the pace at which individuals communicate, exchange information, and collaborate across social digital networks. This change requires, to a great extent, competences in processing various forms of information, interacting with others, and solving problems which may be distributed across several contexts (e.g., school, workplace, home and social networks). These competences have been labeled as “21st century skills” and comprise communication, collaboration, critical thinking, information literacy, reasoning, creativity, metacognition, problem solving, and argumentation – among others (Binkley et al., 2012; P21, 2012; Suto, 2013; Voogt & Roblin, 2010). The Organization for Economic Co-operation and Development (OECD) defined 21st century skills as “skills and competencies young people will be required to have in order to be effective workers and citizens in the knowledge society of the 21st century” (Ananiadou & Claro, 2009, p. 8). Skills related to communication, collaboration, and problem solving have gained considerable attention (Greiff, Niepel, & Wüstenberg, 2015). For instance, within the Assessment and Teaching of 21st Century skills (ATC21s) project, a comprehensive framework of 21st century skills comprising ten skillsets was developed (Griffin et al., 2012).

This is a pre-print version.
DOI to published version: 10.1016/j.chb.2017.08.007

These skillsets were grouped in four domains: Ways of thinking (*Creativity and innovation*; *Critical thinking*, *Problem solving*, *Decision making; Learning to learn, Metacognition*); Ways of working (*Communication*; *Collaboration*); Tools for working (*Information literacy*; *ICT literacy*); Living in the world (*Citizenship; Life and career*; *Personal and social responsibility*). Hesse et al. (2015) argued that these skillsets are multifaceted, multidimensional, and complex. Although researchers have pointed to the necessity of valid assessments to inform teachers and stakeholders about the success and status of incorporation of 21st century skills in instruction (Ras et al., 2014), only limited research has been conducted, particularly on students' collaborative problem solving skills (Quellmalz, 2009).

Collaborative Problem Solving (ColPS)

ColPS is a construct that reflects both collaboration *and* problem solving. Roschelle and Teasley (1995) defined *collaboration* as "coordinated, synchronous activity that is the result of a continued attempt to construct and maintain a shared conception of a problem" (p. 70). *Problem solving* has been defined as "cognitive processing directed at transforming a given situation into a goal situation when no obvious method of solution is available" (Mayer, 1990, p. 284). Weaving the two constructs together, the Programme for International Student Assessment (PISA) defined *ColPS* as "the capacity of an individual to effectively engage in a process whereby two or more agents attempt to solve a problem by sharing the understanding and effort required to come to a solution and pooling their knowledge, skills and efforts to reach that solution" (OECD, 2013, p. 6). The international project ATC21S conceptualized *ColPS* slightly differently as "the abilities to recognize the point of view of other persons in a group; contribute knowledge, experience, and expertise in a constructive way; identify the need for contributions and how to manage them; recognize structure and procedure involved in resolving a problem; and as a member of the group, build and develop group knowledge and understanding" (Griffin, Care, & McGaw, 2012, p. 7). This project outlined a framework of ColPS by distinguishing between two broad skillsets: *Social skills* and *Cognitive skills*.

This is a pre-print version.
DOI to published version: 10.1016/j.chb.2017.08.007

Social skills relate to the “collaborative” part of ColPS, while cognitive skills relate to the “problem solving” part. Hesse and colleagues (2015) argued that the social skills concern managing participants (including oneself), whereas cognitive skills concern managing the task at hand. Along these lines, the *social skillset* comprises three categories of indicators: participation, perspective taking, and social regulation; the *cognitive skillset* comprises two categories: task regulation, and learning and knowledge building. Table 1 details these five categories and presents the indicators of each. Hesse et al. (2015) argued that even though the skillsets and related indicators describe collaborative problem solving, “it is not the case that collaborative problem solving *skills* can be easily mapped to the different stages. Rather, many skills cut across several problem solving stages” (p. 41). This complexity of ColPS processes challenges the development and evaluation of valid assessments.

Computer-Based Assessments of ColPS

As mentioned earlier, students’ ColPS skills have scarcely been measured, and especially not in standardized testing environments. However, a limited number of approaches have been taken. ColPS was for instance incorporated in the Programme for International Student Assessment (PISA) 2015 as a cross-curricular competence (OECD, 2013). It was measured by tasks in which students had to interact and solve problems with a virtual agent (i.e., a computer-simulated avatar). This approach of assessing ColPS using student-agent interactions has several advantages in standardized testing. For instance, all students receive the same stimulus while solving the tasks, and the scoring procedures are automated. However, one of the drawbacks of collaborating with an agent is that the students are not provided with an authentic experience. The authenticity of the problem solving situation may resemble students’ real-life experiences and therefore enhance the transferability of ColPS skills to similar experiences. Besides, interaction with peers may increase students’ motivation and engagement with the assignment. More importantly, the interaction with an agent cannot entirely capture the dynamics of interaction between humans.

This is a pre-print version.
DOI to published version: 10.1016/j.chb.2017.08.007

In another study aiming at describing the processes of ColPS, Bernard and colleagues (2009) studied three forms of interaction conditions (e.g., student-teacher, student-student, and student-content) and determined their effects on student learning. They found that all three interactions were positively and significantly related to learning; nevertheless, the student-student interaction was the most important among the three. These results were supported by Schmid et al. (2014) who compared student-student interactions in postsecondary classrooms between groups with varying degrees of technology use. They discovered that the presence of technology improved the effectiveness of interactions among students for enhancing student achievement. Along these lines, a recent meta-analysis revealed that collaborative qualities – when purposefully added to technology-supported student-student interaction – contribute substantially to learning (Borokhovski, Bernard, Tamim, Schmid, & Sokolovskaya, 2015). The authors further concluded that “instructional design that supports pedagogically sound planning and the implementation of collaborative activities […] [is] the clear winner” (p. 21). These findings suggest that collaboration between students has an impact on their achievement and motivation; yet little is said about the content of the collaborative projects.

Think-Aloud Protocols (TAP)

The TAP methodology has been used for decades to increase the comprehension of human thoughts and actions (Duncker, 1945; Ericsson & Simon, 1980). Studies using TAP focus on the participants’ “think aloud”, that is, their verbalization of their thinking while taking a task. The TAP methodology is increasingly used in educational research as it provides rich and in-depth data on thinking processes (Pressley & Afflerbach, 1995). The use of TAP has also been emphasized in usability testing (often labeled as cognitive labs), product development, and design, and was first proposed by Clayton Lewis (1982). One reason for the increased use of TAP may lie in the fact that this methodology uncovers aspects of students’ thinking processes that may not be captured by performance-based tasks; TAPs consequently

This is a pre-print version.
DOI to published version: 10.1016/j.chb.2017.08.007

contribute to crafting a validity argument (Wilson, 2005). Nevertheless, Ericsson and Simon (1993) pointed out that concurrent utterances are often incoherent and may therefore provide perplexing data. Furthermore, Pressley and Afflerbach (1995) emphasized the richness and variability of language surfacing in TAP, which may also reflect liabilities of the spoken language, and challenge the validity of the TAP analysis. Still, TAP is a powerful data collection methodology that provides data on cognitive processes, access to reasoning procedures underlying cognition, and students' decision making (Greene et al., 2014; Pressley & Afflerbach, 1995). Scholars further highlighted that verbalizations that take place concurrently with cognitive processes are to a great extent independent of participants' interpretations, and thus not affected by perception (Ericsson & Simon, 1993; Van Someren, Barnard, & Sandberg, 1994). Nevertheless, collecting TAP data is both labor-intense and time-consuming (Ericsson & Simon, 1993). Even though only small sample sizes are surveyed, the resultant data sets might become very rich and large and still provide valid information (Johnstone, Altman, & Moore, 2011). Nielson (1994) suggested that a sample size of five participants may already yield sufficient information about problem solving behavior. This brings to attention that participants need to be selected purposefully; emphasizing the diversity among participants is considered important.

Finally, we notice that the TAP methodology is by no means considered equivalent to all other assessment evaluation procedures (e.g., expert reviews, psychometric analyses), but may provide otherwise unexploited information about the test design and students' performance (Kopriva, 2001).

The Present Study

In light of the research presented, the approaches to develop and empirically investigate ColPS tasks for educational assessment are scarcely described. In particular, a lack of tasks that can be utilized by teachers, for instance, to generate formative feedback, identify students' ability levels with regard to ColPS, and for supporting instruction has been

This is a pre-print version.
DOI to published version: 10.1016/j.chb.2017.08.007

identified (Ercikan & Oliveri, 2016). Ercikan and Oliveri further argue that the complex nature of constructs within 21st century skills frameworks require the investigation of quality evidence to go beyond traditional analyses, and suggest that empirical evidence based on students' reporting of their thinking and response processes is vital.

The ATC21s project has dealt with some of the issues and developed the *Learning in Digital Networks* (LDN-ICT) test – a test facilitating synchronous collaboration between students (Griffin & Care, 2015). This test was supplemented by a newly developed ColPS task, which is the primary focus of the present study.

The aim of this study was two-fold: First, we sought to develop a ColPS task that measures the social and cognitive skillsets, as described in the underlying framework (Hesse et al., 2015). Second, on the basis of this ColPS task, we were aimed at evaluating in-depth information about student-student interactions in a digital environment. These aims were guided by the following, overarching research question:

To what extent does the ColPS task facilitate interaction and problem solving between the students in a group while working on a performance-based assessment?

Method

The LDN-ICT test

The LDN-ICT test was developed in the context of the ATC21s project. As part of the development and validation process, a panel of teachers and national project managers reviewed the initial tasks considering (1) the possibility for applying the tasks for both instructional and assessment purposes; (2) the extent to which the tasks discriminate between high and low achievers; (3) whether the tasks engage and motivate the students; and (4) whether the tasks provide sufficient information for scoring and reporting (Griffin & Care, 2015). The LDN-ICT test is constructed to measure students' learning in digital networks – a skillset across several competence areas (e.g., ICT literacy, communication, collaboration,

This is a pre-print version.
DOI to published version: 10.1016/j.chb.2017.08.007

citizenship, ColPS, and technical skills; Wilson & Scalise, 2015; Wilson, Scalise, & Gochyyev, 2015). This test has been further developed and translated to adequately fit the Norwegian language and school culture. LDN-ICT contains three scenarios (i.e., modules or testlets), each of which includes a number of tasks and relate to three different contexts (Scenario 1: *Arctic Trek*, Mathematics and Natural Science; Scenario 2: *Human Legacy*, Social science and Arts; and Scenario 3: *Second Language Chat*, Language). In the present study, the *Human Legacy* scenario was adapted, and a novel task (*CoSketch*) was developed in order to assess and evaluate students' ColPS competences.

The *Human Legacy* scenario

The *Human Legacy*[1] scenario was framed as part of a poetry work unit, in which students were supposed to read and analyse well-known poems (Wilson & Scalise, 2015). Attempts were made to keep the tasks close to authentic classroom situations. For instance, in a typical classroom context, the teacher might find it difficult to make the students express the moods and meanings of a poem. Moreover, the students might wait to hear the teacher's understandings of the poem first, and then agree with it. In order to help students to formulate and structure their ideas about the poem they were given different tasks.

The entire set of tasks in the LDN-ICT test was translated into Norwegian, and further revisions were implemented. The aim was to keep the translated version as close as possible to the original test. Nevertheless, the translation of the *Human legacy* scenario required larger changes and replacements of some tasks. For instance, the English poems were replaced by Norwegian poems Norwegian students were familiar with in their classroom settings. Furthermore, authentic activities that typically occur in classrooms were incorporated in the tasks. For example, the students were asked to express their interpretation of the moods and meanings of the poem, and whether the YouTube-video they watched was an adequate

[1] Note that Wilson and Scalise (2015) labeled this scenario *Webspiration*. We chose the labeling *Human legacy* instead, because *Webspiration* refers to the software used in the original test which was not available for the translated test.

This is a pre-print version.
DOI to published version: 10.1016/j.chb.2017.08.007

interpretation of the poem. As mentioned earlier, the revised and adapted LDN-ICT Human Legacy scenario was supplemented by a newly developed task, *CoSketch*, with the aim of investigating students' problem solving competences while communicating and collaborating with their peers.

Construction of a new ColPS task, *CoSketch*

In this study, we designed and investigated a task which requires students to solve a problem in collaboration within an assessment. The aim was to construct a fairly open, real-world task without too many restrictions. The task was embedded in a performance-based online assessment with open access to the Internet, and chat software was used in which students could communicate synchronously without further restrictions, as opposed to phrase-based chats (Hsieh, & O'Neil, 2002; Rosen, & Foltz, 2014).

The ColPS task *CoSketch* consists of integrated multiple technologies (Jeong & Hmelo-Silver, 2012) and offers shared software for both communication, and creation of a drawing. The integrated software comprises a substantial advantage for the students, as they can interact with each other and create the product they are asked to (i.e., a drawing) for solving the task at the same time. Hence, students can engage in the activities simultaneously and share the same information and processes. In comparison to the remaining LDN-ICT tasks which also include multiple, yet not integrated technologies, we assume that sharing and interacting in a joint assignment, facilitated by fully shared software, may positively affect and stimulate the student-student interaction. This integrative software feature may lessen the burden of shifting between the different technologies (e.g., chat, the test environment, and the software for creating a drawing or painting).

A number of considerations guided the construction of the ColPS task. In a research report on collaboration, Lai (2011) emphasized that the purpose of measuring students' collaboration should be steering the construction of the tasks, and emphasis needs to be put on different aspects of collaboration. Webb (1995) proposed a framework for designing group-

This is a pre-print version.
DOI to published version: 10.1016/j.chb.2017.08.007

based assessments, highlighting the importance of the *process* rather than the *product* of collaboration if the intended purpose is to measure students' ability to collaborate. Moreover, the construction of the tasks was informed by the conceptual framework of ColPS that was developed within the ATC21S project (Hesse et al., 2015). Because this framework distinguishes between the two skillsets, social skills and cognitive skills (Table 1), each of these perspectives were integrated in the task. In order to circumvent a heavy reliance on mathematics, science, or other subject knowledge, we chose relatively easy contents for this task.

CoSketch followed a multi-stage design, in which students had to solve sub-tasks individually before and after the collaborative sub-task. Webb (1995) argued that this is an ideal approach for understanding participants' ability to learn from the collaboration. The individual tasks required the students to access an external webpage (Figure 1), on which students could read the poem and answer questions about it; furthermore, they were able to watch a YouTube video about the poem and answer questions related to this video. One of the questions referred to whether the students believed that the video showed a good interpretation of the poem. Next, students were given a task in which they were asked to develop a mind map about the poem. They were supposed to add "moods" and "meanings" of the poem and connect the different concepts they had identified (Figure 2). These tasks were created in order to help students reflect and develop their *own* ideas about the poem. The subsequent task required students to meet their group by accessing a webpage link to the online software, *CoSketch* (Figure 3). This software consists of a drawing tool and an embedded chat. Students were asked to sketch a drawing together in groups of four students (Figure 4). The resultant drawing should express their interpretation of the poem they had just read and watched a video about. Since *CoSketch* is shared software and only allows for one drawing at a time, the students had to draw together and collaborate by, for instance, drawing, erasing or revising the shared sketch. At the end of the task, students were supposed to save

This is a pre-print version.
DOI to published version: 10.1016/j.chb.2017.08.007

and upload the collaboratively sketched drawing along with the group conversation (i.e., the chat).

After the ColPS task, the students were again given individual tasks related to the poem. Students were asked to sort questions about the poem from least to most useful (Figure 5), and next to drag cards including claims about the poem into the correct category (Figure 6). In the end, the students were asked to evaluate the collaboration with their peers and indicate to what extent the collaboration fostered their understanding of the poem. The Norwegian poem *Regn Regn* written by Sigbjørn Obstfelder was used in the test. It is a short and simple poem about rain, and on rhyme.

Sample

TAP data of twelve Norwegian students in grade 9 and the first year of upper secondary school were collected between February and September 2015. A number of schools were contacted for a larger pilot of the LDN-ICT test. The respondents for the TAP study were selected from three different schools, and two different classes within each school. The students were selected by their teachers according to the maximum variation criteria (Onwuegbuzie & Collins, 2007). The aim was to have a variety of students (e.g., gender, ethnicity, and educational achievement) in each group, to investigate whether the ColPS task facilitates collaboration between the different individuals in a group. Previous studies have focused on pairing students with similar abilities (Rosen, & Tager, 2013), which may decrease the free-rider effect (Bossert, 1988; Salomon, & Globerson, 1989). However, investigating the collaboration between students with different abilities decreases the level of effort – such that the teacher or assessment administrator does not need to identify and pair students with equal abilities. Moreover, a varied group is more in line with what happens in both education and employment – people seldom get to choose whom to collaborate with. The group of older students was chosen from a validity perspective; we sought to investigate whether the constructed task is appropriate for ninth graders or older students.

This is a pre-print version.
DOI to published version: 10.1016/j.chb.2017.08.007

The four students from each school formed one group which virtually collaborated during the test. Thus, the data comprised three groups (namely A, B and C). As shown in table 3, Groups A and B consisted of students in grade 9; group C consisted of students in their first year of upper secondary school. This group initially consisted of four students but due to technical issues one student did not manage to participate in the collaborative task, and was excluded from the dataset. Each group consisted of boys and girls, and a variety of high to low achieving students (see Table 3).

Data collection and assessment procedure

The field work began with a meeting between the principal investigator (PI) and the students at each school. The PI explained the test and the TAP procedures, and demonstrated the process of thinking aloud. Furthermore, the four students in each school received individual usernames and passwords for entering the online test using laptops. After the introduction, students were placed in four different rooms, and logged into the test to access the scenario *Human legacy*. The software *CamStacia*[2] was used to record all activities on the students' computer screen and their audio (i.e., thinking aloud). The PI visited the four students to answer questions, and made sure they followed the thinking-aloud procedure. The time students could spend on the Human legacy scenario was limited to 45 minutes; they were able to monitor the number of tasks and how many they had left on the bottom of the test window (see figure 1).

In this study, TAP was utilized for a number of reasons: (1) to examine the usability and authenticity of the assessment tool, with a main focus on the revised and newly developed tasks; (2) to uncover a more in-depth view of the processes related to collaboration and solving tasks collaboratively in an assessment situation beyond the evaluation of the correctness of students' responses to the tasks; and (3) to obtain information from the TAP analyses in order to improve the assessment tool.

[2] https://www.techsmith.com/

This is a pre-print version.
DOI to published version: 10.1016/j.chb.2017.08.007

Coding of TAP data

The TAP data comprised video material of students' screens (e.g., actions on the computer and within the test environment) and audio of their thinking alouds. The software *InterAct*[3] was used for coding and data analysis. The PI and a student assistant developed a coding scheme in an iterative process. First, the most relevant codes were identified on the basis of the underlying framework of ColPS (Hesse, 2015; Table 1). Second, one TAP file was coded using these indicators, and further indicators were identified given the observed interactions between the students in a group while solving the ColPS task. These include both *event* and *sequence* codes, each of which can be assigned to a certain time and frame in the data file. The resultant coding scheme distinguished between five main categories: Time, pre-individual portion (initial tasks), student-student interactions, post-individual portion (follow-up tasks), incidents, and help. Detailed descriptions of the indicators in each category are described in Table 2. The category *Time* consisted of six indicators for the time each student spent on the individual tasks in the beginning, the total time spent on the ColPS task, and the time spent on the ColPS task without anyone to collaborate with (i.e., when one student was completely on his or her own in the ColPS task environment). Moreover, time spent on post-individual tasks and non-task behavior was calculated. The category *Pre-individual portion* consisted of eight indicators which were related to the individual tasks and processes students had to go through before they entered the ColPS task. These were coded as events, and were used as indicators of whether the students had been familiarized with the poem in different ways, so the team members could bring individual perceptions and opinions to the ColPS task (Nihalani et al., 2012). In addition, each of the tasks in which the students had to add a response were scored with partial credits (i.e., 0 = did not solve it or wrong answer; 1 = below average (i.e., somewhat correct answer); 2 = average (partly correct answer); 3 = above average (correct answer); see Appendix A for an overview of the scoring of each task). The

[3] https://www.mangold-international.com/en/

category *student-student interactions* consisted of three sub-categories (Table 2) which were derived from the two main skillsets *Social skills* and *Cognitive skills* in the underlying ColPS framework (Hesse et al., 2015; Table 1). These three categories included the main indicators of students' ColPS competence: students' role taking during the collaboration (i.e., drawer, observer, writer, and saboteur; see Table 2), students' Cognitive skills (GoalSetting, ClarifyProcess, TimeManagement, Draws, and Task_discussion; see Table 2) and Social skills (introduceOneSelf, SeeksToKnowWhosIsInChat, SeeksHelpFromOthers, GivesSupportToOthers, Social_discussion, and OffTaskBehavior: see Table 2). These indicators are further aligned with the ColPS framework in Table 1, and labeled *Coding rubric* and placed next to the corresponding categories in the framework. As can be seen from Table 1, all elements except *Audience awareness* were assigned to an indicator in the coding scheme. Some of the codes were assigned to more than one element in the framework. This was due to the fact that some elements in the framework were challenging to disentangle. Moreover, the ColPS task was not entirely constructed as an ill-structured problem (e.g., tasks that cannot be solved by recalling facts or be solved by a single, competent group member; Hesse et al., 2015; Rosen, & Tager, 2013) with regard to the time, effort, and the cognitive demands required for solving it. Hence, a more fine-grained coding to identify these nuances could not be applied. However, the task was challenging and partly ill-structured because it required the students to collaborate to solve the problem by using the shared software.

The two indicators *task related discussion and social discussion* were added to the coding scheme to gauge students' contributions to the communication related to either the task (labeled as *taskDiscussion;* see Tables 1 and 2) or regarding other tasks in the test, or issues related to the test situation (labeled as *socialDiscussion;* see Tables 1 and 2). The category *socialDiscussion* was used for capturing students' communication regarding other issues than those strictly related to solving the task (still related to the test), and should not be confused with non-task behavior which was used for coding students' behavior related to

This is a pre-print version.
DOI to published version: 10.1016/j.chb.2017.08.007

other issues than the test. Note that the indicators in the ATC21S ColPS framework are defined at three levels (e.g., low, middle, high; see Hesse et al., 2015, p. 43). However, the indicators were not differentiated in our coding rubric as we wanted to seek the processes in students' interaction and not the detailed level of each indicator.

The indicators *Incidents and Help* (i.e., Technical_Issues, Help_tech and Help_task: see Table 2) were used each time the student encountered technical problems and asked for or received help related to solving the tasks or technical issues.

Finally, the category *post-individual portion* which consisted of seven indicators was coded and scored (see Appendix A). The end-product (i.e., the drawing) was evaluated and scored according to how well it aligned with the poem. In addition, for each student four overall categories were evaluated (see Appendix A). These comprised indicators related to the students' collaboration skills (evalCollab), whether he or she collaborated further in tasks other than *CoSketch* (FurtherCollab).

One file was coded at a time, and each time the indicator appeared for that particular student it was coded as an event (occurrence) or time lap. The coded data files of each student were secondly grouped together. In-depth analyses were conducted on each group file. Moreover, to assure the quality and consistency of the coding, the PI and a research assistant coded two full data sets together and discussed and refined the coding scheme and the coding. Secondly, the remaining data files were divided and coded by the PI and the research assistant. To ensure the quality of these processes, insecurity and disagreements related to the coding and analysis were resolved through discussions within the research team.

Results

In order to address the research question, *To what extent does the ColPS task facilitate interaction and problem solving between the students in a group within an assessment*, we started with a vertical (within-case) analysis by comparing each student's contribution to the

This is a pre-print version.
DOI to published version: 10.1016/j.chb.2017.08.007

ColPS task within the group (Miles & Huberman, 1994). Second, the results of the vertical analysis of each group were submitted to a horizontal analysis (i.e., cross-group comparisons; Miles & Huberman, 1994), in which the three groups were compared for investigating similarities and differences.

Vertical (within-case) analysis

Group A. This group consists of 3 male students and a female student who differed in their levels of school achievement (see Table 3: column labeled *Grades*). Table 3 indicates that the high-achieving students in group A also wrote the most during the collaborative task (i.e., column labeled: *Text*) and used SMS-language the least (Table 3: column *SMS*). Students' collaboration was evaluated as low (Table 3: column *EvalCollab*), and they made a sketch which represented an average level of achievement (Table 3: column *Drawing*).

Time. The four students in this group spent different amounts of time on the collaboration task within a range between 12 and 23 minutes (Table 3: Column *Time*); they spent roughly the same time in the *CoSketch* environment. During the 23 minutes in in the *CoSketch* environment, at least two students were present at almost any time point. The time profile of the students in group A is shown in Figure 7a.

Student-student interactions. Figure 7b shows the actions taken by the students in the ColPS task (e.g., the codes: introduce him- or herself; seeksToKnowWhoIsInChat; seeksHelp; SupportsOthers; TaskSiscussion; SocialDdiscussion; Draws). In group A, student 01 has a limited contribution to the collaborative part of the task (Figure 7b). He does neither engage much in the task discussion nor does he participate in the social discussion or any other interaction in order to solve the task collaboratively. However, the student contributes by drawing, and takes on the role as an observer. This was indicated by (a) his appearance in the chat room, and (b) the observation that he did not write much but sketched in line with what the other group members suggested. Students02-04 were equally engaged in solving the task, and contributed all to the task- and social discussions. All contributed to the drawing, but

This is a pre-print version.
DOI to published version: 10.1016/j.chb.2017.08.007

Student03 drew the most while Student04 was more engaged in the task discussion —he also sought help from the others more frequently than the others (Figure 7b). Students02 and 03 tried to clarify the process to come to a solution by for instance analyzing and rephrasing the task, sharing responsibility for the different parts of the process to solve the problem, and asking questions related to time and capabilities of the group members (see also Table 1). Student03 contributed the most to the sketch and took the role as drawer. He also tried to identify the other students' identities in the chat (Figure 7b); this student wrote most in the chat (Table 3:Text).

Most of the students were positive to the test and especially to the collaborative task. Student04 expressed her emotions related to the task very enthusiastically and following is an excerpt from her think-aloud protocol, including the time stamp..

14.27: "Now we will draw - what, what the poem is about"

17.44: "Now we will draw some skies and such concerning what this poem is about. We are a group of four that will collaborate to make a drawing."

18.25: "And everybody draws what they think it is about. Skies and rain and, so everybody is collaborating."

25.18: "It was fun to see what the others answered. Not fun, just what they were thinking in a way"

25.30: "The collaboration helped me to better understand. Yes it did so a little actually"

25.55: "Explain how your understanding of the poem changed through the collaboration. We did manage to draw and show what we understood about the poem and through that I understood better that it actually is not only about rain, rain all the time".

Group A's interaction processes are shown in Figure 7c. The vertical axis in the figure indicates the time line and the horizontal axis the *interactions*. Students in this group are represented by circles with different colors, and the small arrows indicate the direction of the processes of their actions within the different time categories. As Figure 7c shows, Student01 (i.e., white circles) and Student04 (i.e., blue circles) interacted the most, while the other two students in the group contributed by drawing.

This is a pre-print version.
DOI to published version: 10.1016/j.chb.2017.08.007

Overall, three of the students were mostly collaborating and solved the task successfully. All students contributed and all codes in the underlying framework were applied..

Group B consisted of 2 male and 2 female students with differing levels of school achievement (Table 3). Students of this group had better school grades than students in the other groups (Table 3: column *Grades*). As shown in Table 3, each student within this group contributed considerably by writing in the chat, and little SMS-language was used. Moreover, students' collaboration was evaluated as average (Table 3: column *EvalCollab*), and the sketch they drew was of low quality (Table 3: column *Drawing*). This group collaborated further on the post-individual tasks, even after they had finished the the collaborative task (Table 3: column *FurtherColl*).

Time. The four students in this group spent differed in the time they had spent on the collaboration task within a range between 16 and 41 minutes (Table 3). For this group, two or more students were in the *CoSketch* environment for around 27 minutes. Student05 accessed the collaboration software around 3 minutes after logging on to the test – this observation explains why she spent 41 minutes in the collaborative task. An overview of the time profile of the students in group B is shown in Figure 8a.

Student-student interactions. As shown in Figure 8b, Student05 and Student07 are equally active and contribute equally to the task-related discussions and the actual process of drawing. Student06 contributes overall the least to the discussion in the chat compared to the other group members. Yet, he draws most of the sketch and therefore was assigned the role *drawer*. Student08 in general contributes more to the overall interaction and leads the task-related discussion, seeks help from the others in the group, and contributes by drawing. This student was assigned the role *leader* since she took on the responsibility to drive the process of understanding the task, discuss with the group towards a solution.

This is a pre-print version.
DOI to published version: 10.1016/j.chb.2017.08.007

Both Student07 and Student08 struggled with uploading the chat and asked the others in the group for help concerning how to save the chat/conversation. A transcript from Student07s' TAP illustrates this:

18.20: "Since it was one is one and two is two I think it is enough with two persons so I might create the water a little smaller"
18.45. "Oh this worked well actually"
22.50: "How can one save the chat? I can just copy the whole thing, but.."
26.05: "It is fine. We can rather continue, and then we can go back later instead if it is possible. Yes, it is possible."
38.19: "But I haven't compared the answers with the others"
43.46: "Should I ask or should I not ask?"
44.13: (writes in the chat) "Task 7? What do you think about it?"
46.08: (writes in the chat) "Did we really collaborate?"
48.53: "It was valuable to compare my results with the others. I do not know exactly what to say because we…"

Moreover, Student08 was quite skeptical to the others in the beginning when she entered the chat. The following excerpt from her TAP exemplifies this:

16.13. "Now I am in the chat-thingi because it was blinking all the time, but I have to hurry to get there"
17.30: "I have no clue who this is and it is really scary»
17.41: "I am actually thinking quite a lot while I am doing this"
18.05: "Wait, who is this? Is it Karen? I want to ask who it is»
21.08: "This is scary. Someone asked me who I am and I don't know who that is"

After she got to know who the other person in the chatroom was, she started to draw, and she got confused when she realized the others could erase her drawing.

24.45: "He can't erase mine. Don't erase my drawing. What is going on now? Well I will have to draw again. I will not give up. Such things irritate me. He can erase his things, but he cannot erase my things"
27.02: "No, no, you are not supposed to draw that. Now I will paint. Now it is my turn."

In the end, she reflected on the collaboration and wrote:

45.13: (writes in the chat) "I think the collaboration did not have a large influence, but it gives other viewpoints about the poem. Which is good".

Group Bs' interaction process is shown in Figure 8c. The figure shows that the students in Group B spent most of the time in the ColPS task together, and interacted throughout the entire task. The density in the figure suggests a high level of interaction between the group members. The figure also illustrates that two of the students (Student07 and Student06) continued collaborating after they had finished the ColPS task.

This is a pre-print version.
DOI to published version: 10.1016/j.chb.2017.08.007

Overall, this group's collaboration processes are characterized by a high level of interaction, and all students in the group took part actively in solving the task and engaging in discussions. They even collaborated during the post-individual part of the assessment. Moreover, all codes were applied, and hence all elements of the underlying framework could be identified during the ColPS task.

Group C. This group comprised two female students and a male student (Table 3). The students were in their first year of upper secondary school (i.e., one year older than students in the other two groups), and differed in their school performance (see Table 3: column *Grades*). All three students wrote large amounts of text and differed in the extent to which they used SMS-language (see Table 3). Their collaboration was evaluated as average (Table 3: column *EvalCollab*), and they made a sketch which represented an intermediate level of achievement (Table 3: column *Task*).

Time. The three students in this group spent different amounts of time on the collaboration task within a range between 23 and 43 minutes (Table 3). Student09 accessed the collaboration software quite early after logging on the test; this explains why she had spent 43 minutes in the joint task. She was also solving the individual tasks while she had accessed the joint task. For this group, two or more students were in the *CoSketch* environment for around 27 minutes. An overview of the time profile of the students in group C is shown in Figure 9a.

Student-student interactions. As shown in Figure 9b, Student09 took most of the responsibility to solve the task in this group. This student was heavily engaged in the task-related discussions, the drawing, supporting the others, and clarifying the process. Student09 also engaged in the social discussion, but less than the other two students in the group. Given the overall performance of Student09, she was assigned the role *leader*. Student10 and Student11 contributed equally to the drawing activities and engaged in the task-related discussions. Student11 aimed at clarifying the processes to come to a solution and sought help

This is a pre-print version.
DOI to published version: 10.1016/j.chb.2017.08.007

from the others in the group. Student10 was the only student who took on the counter-productive role as a “saboteur”. He erased the drawings several times, but he was engaged in the task-related discussions. This groups’ sketch scored at a high level (Table 3). This was the only group that was concerned with keeping track of the time spent on the test and the collaborative task.

Student09 and Student11 encountered technical problems and asked for help from the test administrator to solve them. These problems were related to accessing the YouTube video, and saving and uploading the chat. Because these basic skills were part of the test, the test administrator could not provide help. Furthermore, Student10 showed negative behavior by erasing the groups drawing over and over, and letting the others think he was another student than he actually was. However, he changed his behavior in the end, and let the group complete the task so that they were able to submit the final product. Student09, in fact, resolved Student10’s counter-productive behavior by pushing the “undo” button in the *CoSketch* software several times. This button helped bringing back the last version of the drawing. She wrote a lot in the chat and also sketched most of the group’s drawing. For instance, she started off writing what they can draw in the chat:

06.05: (writes in the chat) “We can for instance draw two kids that jump and it rains at the same time or something like that”

When she realized that the others were not in the chat, she went back to the individual tasks and said:

08.00: “Just now Sarah left the chat and now I don’t know what we should draw because I did not understand the task. Or I understood it, but, I don’t know where the others in the group are and it is a little difficult to make it alone”

After 5 minutes the others are back in the chat, and she continued:

13.10: “Now the others are back in the chat so now I hope I will get done with the drawing”
20.30: (writes in the chat) “Let’s draw sand or something”
28.13: (writes in the chat) “Damn, where is it”. (Says loudly) “the drawing has disappeared because the other student pushed the “clear all” button”.

This is a pre-print version.
DOI to published version: 10.1016/j.chb.2017.08.007

After this and the struggle the group has with Student10 sabotaging the drawing, they managed to finish the sketch. But Student09 did not manage the task which required to save and upload the chat:

40.55: “I did not understand how to upload the chat so I skipped it, because the others are not replying either”

However, in the end she seemed content:

45.51: “Explain how your understanding of the poem changed through the collaboration. It was in fact quite fun to draw our understanding of the poem together. I understood that it is about kids that love rain, and what they did while it was raining”.

Figure 9c visualizes the process of the students’ interactions in group C. Following the coding of the timeline, it illustrates that Student10 spent much time alone in the chat, while Student09 and Student11 collaborated and spent much of their time together in the ColPS task. It is also evident from the figure that Student09 engaged more in solving the task compared to Student11.

Overall, this group’s collaboration processes are characterized by especially one student taking the lead. The second student also contributed to solve the task. And even with a member that was sabotaging, they managed to solve the task well. Once again, all codes were applied.

Horizontal (cross-group) analysis

In general, the students’ conversations were related to either the ColPS task or one of the other tasks in the assessment. Hence, there was little communication about issues other than those related to the *Human legacy* scenario. None of the students visited irrelevant or distracting Internet pages during the test (e.g., Facebook or pages related to news, sports, or music). In the focus group interview afterwards, the students explained that they were engaged with the test and did neither think of accessing other pages nor felt they had the time or urge to do so.

This is a pre-print version.
DOI to published version: 10.1016/j.chb.2017.08.007

Figure 10 illustrates the comparisons between the three groups related to their interactions during the ColPS task. Group B discussed the ColPS task the most, while group A had the least task-related discussions. Students in group B sought more help from each other, and also supported each other more than students in the other groups (Figure 10). Figure 10 illustrates the amount of interaction in the groups and shows that, although little interaction was identified in group A, students put significant effort in the actual drawing activity. Despite the fact that one of the students in group C disrupted the collaborative problem solving process, this group scored the highest on the task.

Discussion

The main objectives of the current study were twofold: First, it was aimed at investigating the extent to which a newly developed ColPS task facilitated student-student interaction and problem solving within an assessment. The second aim was to investigate the extent to which the underlying theoretical ColPS framework could be applied to capture primarily the collaborative processes during the ColPS task. These two intertwined aims were explored with the help of students' TAP data that contained the verbalisation of their thinking as well as their actions and performance on the assessment. The key findings of this study suggest that: (1) the main categories *Time* and *Interactions* provide essential information about individual and group behaviour; (2) TAP data provide valuable information about task behaviour – thus, suggestions for the refinement of ColPS tasks are provided; (3) the ColPS framework can be operationalized, although not all indicators might be clear-cut.

Collaboration time

Our results regarding the time students' spent on the collaborative task during the Human legacy scenario show that the different groups varied in the extent to which they exploited the time to collaborate. This variation was most probably due to the fact that students within a group did not manage to access the task at the same time. For instance, in

This is a pre-print version.
DOI to published version: 10.1016/j.chb.2017.08.007

two observations, the students that did not manage to access the ColPS task synchronously (i.e., when other students were in the *CoSketch* environment) and thus missed out the opportunity to interact and collaborate. Our findings consequently indicate that there is a relation between the time spent on the collaborative task and the synchronization of students' access. This, in turn, may further influence the intensity and frequency of student-student interactions. Another line of thinking interprets time on task information as information about the effort students take in order to engage in either collaborative or cognitive activities in ColPS. Moreover, missing out opportunities to collaborate might also be due to the fact that some students may not be willing or open to collaborate digitally. From an assessment perspective, on the one hand, it might be a positive asset that the test design allows the students to enter the collaborative task at different times, such that the students may decide on their own pace of engaging in collaboration. On the other hand, given the individual differences in time on task, this openness of the environment may create situations in which students miss out collaborative activities by design. This may further affect the role-taking of individual students. Students who enter early may have a better chance of understanding the problem and the environment than those who enter later; these students are therefore more likely to take a lead and/or contribute substantially more than their fellow students.

Our findings further inform the design of assessment. For instance, the assessment should be redesigned such that the ColPS task is accessed earlier. Still, the pre-individual portion is important to prepare the students for the ColPS tasks and to stimulate them to have their own opinions and ideas. However, for the *Human legacy* scenario, we realise that there were too many tasks in the beginning which affected the time the students accessed the ColPS task. The students spent different amount of time on these tasks, hence we believe this refinement could help students within groups to compensate for the asynchronous access to the collaborative task. Steering the time students are allowed to spend on each portion of the

This is a pre-print version.
DOI to published version: 10.1016/j.chb.2017.08.007

task – be it the individual or the collaborative task – is an important design consideration for structuring assessments of collaborative problem solving.

Student-student interactions and role-taking

Our results indicate that students took different roles during the collaborative task, despite the fact that this was not explicitly required in the task. Students may have found it necessary to contribute in different ways to solve the task – especially because it would have not be useful if all group members engaged in drawing simultaneously. Whereas our study allowed students to *freely choose* their roles within their group, it might be worthwhile studying situations in which students *are assigned to specific roles*. Systematically varying the degree of choice in role-taking can provide insights into students' role identification and role switching – two essential elements of collaboration skills (REF). From a design perspective, this could either be simply stated in the task information, suggesting that the students should contribute by taking on different roles (e.g., leader, drawer, organizer) but leaving it open which of these roles each individual student would take or have them pre-assigned to specific roles randomly in the beginning of the collaborative task. Moreover, task designers may also encourage students to discuss and share roles as part of the task (e.g., an example of this arrangement in a task is provided in Wilson & Scalise, 2015). From a methodological point of view, think-aloud protocols have proved useful in this study, particularly in order to uncover student-student interactions and student behavior.

- For theory about roles - Hesse and Rosen and Tager

In this test, all students received the same task stimulus (i.e., they watched the same YouTube video and answered the same questions about the poem). For future administration of the task, we may explore the effect of different stimulus on students' collaborative skills. For instance, group members may access and watch different videos about the poem, are asked different questions, or are provided with different pieces of relevant information needed

This is a pre-print version.
DOI to published version: 10.1016/j.chb.2017.08.007

to solve the problem (Jeong & Hmelo-Silver, 2016). These changes in the design of the task may facilitate varying levels of student-student interaction.

Group composition and ability levels

Students with different backgrounds, gender, and ability levels were grouped together to certify that the ColPS task enables the differing group members to interact and engage in creating a solution. In contrast to previous research which paired students with same ability levels and ColPS characteristics (Rosen, & Tager, 2013), the present study performed random pairing of students, because the identification of students' ability levels and further grouping them in the computer-based test environment seemed resource demanding. Moreover, the existing body of literature describing the potential characteristics that might influence students' behaviour in collaborative problem solving environments has not yet reported clear-cut criteria for grouping students (REF). In contrast, pre-assigned roles and responsibilities for specific tasks while solve a problem collaboratively might help to reduce or even circumvent counter-productive behaviour (REF). From a classroom perspective though, it seems challenging for teachers to pair the students on the basis of both the students' social and cognitive skills (Webb, 1995). This is probably an effect of easy-to-use classroom assessments of these skills. Finally, in work life, people often need to collaborate without much knowledge about the others level of skills or knowledge (REF); the random grouping performed in our study might therefore represent a more realistic scenario mimicking real-life situations. This claim, however, should be subject to further testing.

We also believe that teaching students' ColPS requires experiencing working with different combinations of students. These experiences will help them sustain adaptability in changing group settings (Pulakos et al., 2000). Our analyses displayed that the students in all groups were able to contribute to the task, despite the diversity within the group. Even the student who showed counter-productive behavior contributed to the task by drawing and discussing. The fact that collaboration was achieved and counter-productive behavior could

This is a pre-print version.
DOI to published version: 10.1016/j.chb.2017.08.007

be reduced indicates students' willingness to collaborate and their adaptability while interacting with their peers within the group.

In conclusion, the choice as to whether a random or fixed assignment of students to specific roles in collaborative problem solving should be performed depends on the purpose of the task. If embedded in an assessment, it may be useful to assign students to explicit roles, and their performance could be scored in relation to the expectations from these roles. If the goal is to examine role-taking behavior as a part of collaborative skills, tasks should be administered that allow students to select their roles and organize themselves to solve the task. Overall, we believe that role-taking is an essential part of teaching and learning of ColPS.

In our study, it appeared that not all students in a group seemed to engage equally in the ColPS task. As previously mentioned, there may be different reasons for this (e.g., time on task, the roles taken). Another, perhaps obvious reason refers to group size. Even though manipulating group sizes was not the core aim of this study, we suggest exploring the assessment with teams of two and/or three students to identify the optimal group size, because tasks which are less cognitively demanding may benefit from smaller groups. Another advantage of smaller groups is that it may reduce free riding or social loafing (Salomon & Globerson, 1989; Webb, 1995), and may increase equality of participation. Researchers argued that group size and composition should be carefully considered as it may affect group performance (Rosen, & Rimor, 2009; Webb, 1995; Wildman et al., 2012).

Operationalisation of the ColPS framework

The indicators of the social and cognitive skillsets developed for encapsulating the students' ColPS processes were applicable in our study. On the basis of the results of both the vertical and horizontal analysis, we could identify that all the indicators were applied to each group. This finding suggests that the ColPS task itself has features that enable students to

This is a pre-print version.
DOI to published version: 10.1016/j.chb.2017.08.007

communicate, collaborate, and solve the problem. We consequently argue that the ColPS task is useful to engage students in collaborative problem solving.

At the same time, we experienced difficulties related to the development of categories to disentangle some of the elements in the framework. Thus, we encourage further research to investigate the alignment between the structure of the framework and the task design. We hypothesize that more cognitive challenging tasks or ill-structured problems (Lai, 2011) may surface other aspects of the social or cognitive processes described in the framework. The issue related to the operationalisation of the ColPS framework has been stressed by several researchers who emphasize that the inherent structure of the ColPS construct has yet not been fully understood (Care et al., 2016).

Conclusion

Researchers have pointed to the complex nature of the ColPS construct and encouraged the design of effective assessment approaches (Care et al., 2016; Rosen, & Foltz, 2014; Rosen, & Tager, 2013). Accordingly, the present study presented and explored a novel ColPS task. Our results showed that the newly developed task facilitates student-student interaction, thus allowing us to identify and describe collaborative problem solving competences. The analysis showed promising results regarding the task's ability to enable all students to contribute to the collaborative problem solving process in different ways. However, further refinements with respect to collaboration time, role-taking, and performance scoring are suggested in order to improve and further investigate the potential of the task. We believe that, with the increasing use of technology, synchronous ColPS competences are crucial and should be both instructed and assessed within educational systems – our study takes a step towards investigating the construction and application of such tasks.

This is a pre-print version.
DOI to published version: 10.1016/j.chb.2017.08.007

References

Abrami, P. C., Bernard, R. M., Bures, E. M., Borokhovski, E., & Tamim, R. (2011). Interaction in distance education and online learning: Using evidence and theory to improve practice. *Journal of Computing in Higher Education, 23*(2/3), 82-103.

Bernard, R. M., Abrami, P. C., Borokhovski, E., Wade, A., Tamim, R., Surkes, M., & Bethel, E. C. (2009). A meta-analysis of three interaction treatments in distance education. *Review of Educational Research*, *79*(3) 1243-1289. http://dx.doi.org/10.3102/0034654309333844

Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., & Rumble, M. (2012). Defining Twenty-First Century Skills. In P. Griffin, B. McGaw, & E. Care (Eds.), *Assessment and Teaching of 21st Century Skills* (pp. 17-66): Springer Netherlands.

Borokhovski, E., Bernard, R. M., Tamim R. M., Schmid R. F., & Sokolovskaya A. (2015). Technology-Supported Student Interaction in Post-Secondary Education: A Meta-Analysis of Designed Versus Contextual Treatments. Computers & Education, 96, Pp. 15–28. Doi: 10.1016/j.compedu.2015.11.004.

Bossert, S. T. (1988). Cooperative activities in the classroom. *Review of Research in Education, 15*(1988-1989), 225–250.

Care, E., Scoular, C., & Griffin, P. (2016). Assessment of Collaborative Problem Solving in Education Environments, Applied Measurement in Education. DOI: 10.1080/08957347.2016.1209204

Care, E., Griffin, P., Scoular, C., Awwal, N., & Zoanetti, N. (2015). Collaborative problem solving tasks. In P. Griffi n & E. Care (Eds.), *Assessment and teaching of 21st century skills: Methods and approach* (pp. 85–104). Dordrecht: Springer.

Csapó, Ainley, Bennett, Latour, & Law. (2012). Technological Issues for Computer-Based Assessment . In Griffin et al. (eds.), *Assessment and Teaching of 21st Century Skills.* DOI 10.1007/978-94-007-2324-5_4, © Springer Science+Business Media B.V.

Collaborative for Academic, Social, and Emotional Learning (CASEL). (2003). *Safe and sound: An educational leader's guide to social and emotional learning programs*. Retrieved from www.casel.org, at 23.04.2015

Duncker, K. (1945). On problem-solving. In Dashiell, J. F. (Ed.) *Psychological Monographs* (pp.1–114). Washington, DC: American Psychological Association.

This is a pre-print version.
DOI to published version: 10.1016/j.chb.2017.08.007

Ericsson, K., & Simon, H. (May 1980). *Verbal reports as data. Psychological Review 87 (3): 215–251. Doi: 10.1037/0033-295X.87.3.215*

Ericsson, K., & Simon, H. (1993). *Protocol analysis: Verbal reports as data (2nd edition).* Cambridge, MA: MIT Press.

Fidalgo-Blanco, Á., Sein-Echaluce, M.L., García-Peñalvo, F.J., & Conde, M. Á. (2015). Using Learning Analytics to improve teamwork assessment. Computers in Human Behavior, 47, pp. 149-156. Doi:10.1016/j.chb.2014.11.050

Gallardo-Echenique, E. E., de Oliveira, M. J., Marqués-Molias, L., & Esteve-Mon, F. (2015). Digital Competence in the Knowledge Society. *MERLOT Journal of Online Learning and Teaching*, *11*(1). Retrieved 20.10.2015 from jolt.merlot.org/vol11no1/Gallardo-Echenique_0315.pdf

Greiff, S., Niepel, C., & Wüstenberg, S. (2015). 21st century skills: International advancements and recent developments. *Thinking Skills and Creativity, 18*, 1-3. doi:10.1016/j.tsc.2015.04.007

Griffin, P., Care, E., & McGaw, B. (2012). The Changing Role of Education and Schools. In Griffin, McGaw & Care (eds.), *Assessment and Teaching of 21st Century Skills*. Springer Science+Business Media B.V. DOI 10.1007/978-94-007-2324-5_1.

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

Griffin, P., & Care, E. (2015). The ATC21S method. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills: Methods and approaches* (pp. 3–33). Dordrecht: Springer. ISBN 978-94-017-9394-0

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, Educational Assessment in an Information Age*. doi:10.1007/978-94-017-9395-7_2

Hsieh, I.-L., & O’Neil, H. F., Jr. (2002). Types of feedback in a computer-based collaborative problem solving group task. *Computers in Human Behavior, 18,* 699-715.

Jeong, H., & Hmelo-Silver, C. E. (2016). Seven Affordances of Computer-Supported Collaborative Learning: How to Support Collaborative Learning? How Can Technologies Help?, Educational Psychologist, 51:2, 247-265, DOI:10.1080/00461520.2016.1158654

Jeong, H., & Hmelo-Silver, C. E. (2012). Technology Supports in CSCL. In J. van Aalst, K. Thompson, M. J. Jacobson, & P. Reinmann (Eds.), The future of learning: Proceedings of the 10th International Conference of the Learning Sciences (ICLS 2012)—Volume 1, Full papers (pp. 339–346). Sydney, Australia: Society of the Learning Sciences.

Johnstone, C.J., Altman, J.R. & Moore, M. (2011). Universal design and the use of cognitive labs. In M. Russell and M. Kavanaugh (Eds.). *Assessing students in the margins (pp.425-442).* Charlotte, NC: Information Age Publishing.

Kanuka, H. & Anderson, T. (1999). Using constructivism in technology mediated learning: Constructing order out of the chaos in the literature. *Radical Pedagogy, 1*(2).

Kopriva, R. (2001). *ELL validity research designs for state academic assessments: An outline of five research designs evaluating the validity of large-scale assessments for English language learners and other test takers*. Paper prepared at the Council of Chief State School Officers Meeting, Houston, TX, June 22–23, 2001.

Law, N., Lee, Y., & Yuen, H. K. (2009). The impact of ICT in education policies on teacher practices and student outcomes in Hong Kong. In F. Scheuermann & F. Pedro (Eds.), *Assessing the effects of ICT in education– Indicators, criteria and benchmarks for international comparisons* (pp. 143–164). France: OECD.

Lewis, C. H. (1982). *Using the "Thinking Aloud" Method In Cognitive Interface Design (Technical report). IBM. RC-9265.*

Mayer, R.E. (1990), "Problem solving", in M. W. Eysenck (ed.), The Blackwell Dictionary of Cognitive Psychology, Basil Blackwell, Oxford, pp. 284-288.

Messick, S. (1995). Validity of Psychological Assessment: Validation of Inferences From Persons' Responses and Performance as Scientific Inquiry Into Score Meaning. *American Psychologist, 50*, 741–749.

Miles, M.B., & Huberman, A.M. (1994*). Qualitative data analysis: An expanded sourcebook*. Sage.

National Research Council. (2012). Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century. Committee on Defining Deeper Learning and 21st Century Skills, James W. Pellegrino and Margaret L. Hilton, Editors. Board on Testing and Assessment and Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Nielsen, J. (1994). Estimating the number of subjects needed for a thinking aloud test. International Journal of Human-Computer Studies, 41, 385–397.

Nihalani, P. K., & Robinson, D. H. (2012). "Collaborative versus individual digital assessments," in *Technology-Based Assessments for 21st Century Skills*, eds M. C.

Mayrath, J. Clarke-Midura, D. H. Robinson, and G. Schraw (Charlotte, NC: Information Age Publishing), 325–344.

Onwuegbuzie, A. J., & Collins, K. M. T. (2007). A typology of mixed methods sampling designs in social science research. *The Qualitative Report, 12*(2), 281-316. Retrieved August 31, 2007, from http://www.nova.edu/ssss/QR/QR12-2/onwuegbuzie2.pdf

Ras, E., Krkovic, K., Greiff, S., Tobias, E., & Maquil, V. (2014). Moving towards the assessment of collaborative problem solving skills with a tangible user interface. TOJET: The Turkish Online Journal of Educational Technology, volume 13, issue 4.

Rienties, B., Tempelaar, D., van den Bossche, P., Gijselaers, W., & Segers, M. (2008). *Students' motivations and their contributions to virtual learning*. Presented at the meeting of International Conference of the Learning Sciences, June 2008, Utrecht, the Netherlands.

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

Rosen, Y., & Rimor, R. (2009). Using collaborative database to enhance students' knowledge construction. *Interdisciplinary Journal of E-Learning and Learning Objects, 5,* 187-195.

Rosen, Y., & Tager, M. (2013). Computer-based Assessment of Collaborative Problem-Solving Skills: Human-to-Agent versus Human-to-Human Approach. Pearson. Research Report.

OECD (Organization for Economic Co-Operation and Development). (2013). PISA 2015 Draft Collaborative Problem Solving Framework. Retrieved February 02, 2016, from https://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Collaborative%20Problem%20Solving%20Framework%20.pdf

P21 (Partnership for 21st Century Skills). (2012). *Learn for the 21st century. A report and mile guide for 21st century skills*. Partnership for 21st Century Skills. Retrieved 12.02.2015 from http://www.p21.org/storage/documents/P21_Report.pdf

Petrides, K. V. (2001). *A psychometric investigation into the construct of emotional intelligence*. Doctoral dissertation. (London, University College London).

Petrides, K. V. & Furnham, A. (2003). Trait emotional intelligence: Behavioural validation in two studies of emotion recognition and reactivity to mood induction. *European Journal of Personality*, 17, 39–57.

This is a pre-print version.
DOI to published version: 10.1016/j.chb.2017.08.007

Pressley, M. & Afflerbach, P. (1995). *Verbal Protocols of Reading: the Nature of Constructively Responsive Reading*. Hillsdale NJ: Lawrence Erlbaum Associates.

Quellmalz, E. (2009). Assessing new technological literacies. In F. Scheuermann, & F. Pedro (Eds.), *Assessing the effects of ICT in education– Indicators, criteria and benchmarks for international comparisons* (pp. 143–164). European Union, France: OECD.

Salomon, G. & Globerson, T. (1989). When teams do not function the way they ought to. *International Journal of Educational Research, 13*(1), 89–100.

Schmid, R. F., Bernard, R. M., Borokhovski, E., Tamim, R. M., Abrami, P. C., Surkes, M. A., Wade, C. A., & Woods, J. (2014). The effects of technology use in postsecondary education: A meta-analysis of classroom applications. *Computers & Education, 72*, 271-291. doi:10.1016/j.compedu.2013.11.002

Siddiq, F., Gochyyev, P., & Wilson, M. (2017). Learning in digital networks – ICT literacy: A novel assessment of students' 21st century skills. *Computers & Education,* 109, 11-37. doi:10.1016/j.compedu.2017.01.014

Siddiq, F., Scherer, R., & Tondeur, J. (2016). Teachers' emphasis on developing students' digital information and communication skills (TEDDICS): A new construct in 21st century education. *Computers & Education, 92-93*, 1-14. doi:10.1016/j.compedu.2015.10.006

Siddiq, F., Hatlevik, O. E., Olsen, R. V., Throndsen, I., & Scherer, R. (2016). Taking a future perspective by learning from the past - A systematic review of assessment instruments that aim to measure primary and secondary school students' ICT literacy. *Educational Research Review,*18, 58-84. Doi: 10.1016/j.edurev.2016.05.002

Silva, E. (2009). Measuring skills for 21st-century learning. *Phi Delta Kappa,* 90(9), 630-634.

Suto, I. (2013). 21st Century Skills: Ancient, ubiquitous, enigmatic? *Research Matters.* A Cambridge Assessment Publication. Issue 15, pp: 2-8. Retrieved 05.04.2016 from: http://www.cambridgeassessment.org.uk/images/130437-21st-century-skills-ancient-ubiquitous-enigmatic-.pdf

Van Someren, M. W., Barnard, Y. F., & Sandberg, J. A. C. (1994). *The think-aloud method: A practical guide to modeling cognitive processes*. San Diego, CA: Academic Press Ltd.

von Davier, A. A., & Halpin, P. F. (2013). Collaborative Problem-Solving and the Assessment of Cognitive Skills: Psychometric Considerations. *ETS Research Report Series, 2013*(2), i-36. doi:10.1002/j.2333-8504.2013.tb02348.x

Wang, S.L., & Lin, S. S.J. (2007). The effects of group composition of self-efficacy and collective efficacy on computer-supported collaborative learning. Computers in Human Behavior, 23, pp. 2256 – 2268. doi:10.1016/j.chb.2006.03.005

Webb, N. M. (1995). Group collaboration in assessment: Multiple objectives, processes, and outcomes. *Educational Evaluation and Policy Analysis, 17*(2), 239–261.

Wildman, J. L., Shuffler, M. L., Lazzara, E. H., Fiore, S. M., Burke, C. S., Salas, E., & Garven, S. (2012). Trust development in swift starting action teams: A multilevel framework. *Group& Organization Management*, *37*(2), 138-170.

Wilson, M. (2005): *Constructing Measures: An Item Response Modeling Approach*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Wilson , Bejar , Scalise , Templin , Wiliam , & Irribarra .(2012). Perspectives on Methodological Issues. In P. Griffi n et al. (eds.), *Assessment and Teaching of 21st Century Skills*. DOI 10.1007/978-94-007-2324-5_3, © Springer Science+Business Media B.V.

Wilson, M., & Scalise, K. (2015). Assessment of Learning in Digital Networks. In, Griffin, P., & Care, E. (Eds.), *Assessment and Teaching of 21st Century Skills Volume 2 - Methods & Approaches*. Chapter 3, P. 57-81. Dordrecht: Springer. DOI 10.1007/978-94-017-9395-7_3

Wilson, M., Scalise, K., & Gochyyev, P. (2015). Rethinking ICT literacy: From computer skills to social network settings. Thinking Skills and Creativity, 21st Century Skills: International Advancements and Recent Developments. Volume 18, p. 65-80. Doi: 10.1016/j.tsc.2015.05.001

Zhu, C., Valcke, M., & Schellens, T. (2009). A cross-cultural study of online collaborative learning. Multicultural Education & Technology Journal, Vol. 3 Iss 1 pp. 33 – 46. Doi: http://dx.doi.org/10.1108/17504970910951138

This is a pre-print version.
DOI to published version: 10.1016/j.chb.2017.08.007

Tables

Table 1

ATC21S ColPS framework. The categories and indicators within each skillset; Social skills and Cognitive skills, and the codes from the coding rubric

Social skillset			**Cognitive skillset**		
Element	**Indicator**	**Coding rubric**	**Element**	**Indicator**	**Coding rubric**
Participation			***Task regulation***		
Action	Activity within environment	introduceOneSelf; Sabotage; SeeksToKnowWhosInChat; Non-task behavior	Organises (problem analysis)	Analyses and describes a problem in familiar language	ClarifyProcess
Interaction	Interacting with, prompting and responding to the contributions of others	SeeksHelpFromOthers; GiveSupportToOthers; SocialDiscussion; TaskDiscussion	Sets goals	Sets a clear goal for a task	GoalSetting
Task completion/ perseverance	Undertaking and completing a task or part of a task individually	Draws	Resource management	Manages resources or people to complete a task	ClarifyProcess TimeManagement
Perspective taking			Flexibility and ambiguity	Accepts ambiguous situations	
Adaptive responsiveness	Ignoring, accepting or adapting contributions of others	TaskDiscussion	Collects elements of information	Explores and understands elements of the task	TaskDiscussion GoalSetting
Audience awareness (Mutual modelling)	Awareness of how to adapt behaviour to Increase suitability for others		Systematicity	Implements possible solutions to a problem and monitors progress	TaskDiscussion
Social regulation			***Learning and knowledge building***		
Negotiation	Achieving a resolution or Reaching compromise	ClarifyProcess TaskDiscussion	Relationships (Represents and formulates)	Identifi es connections and patterns between and among elements of knowledge	TaskDiscussion
Self evaluation (Metamemory)	Recognising own strengths and weaknesses	SupportToOthers	Rules: “If … then”	Uses understanding of cause and effect to develop a plan	ClarifyProcess
Transactive memory	Recognising strengths and weaknesses of others	SeeksHelpFromOthers	Hypothesis “what if…” (Reflects and monitors)	Adapts reasoning or course of action as information or circumstances change	TaskDiscussion
Responsibility initiative	Assuming responsibility for ensuring parts of task are completed by the group	TaskDiscussion ClarifyProcess			

DOI to published version: 10.1016/j.chb.2017.08.007

Table 2.

Rubric for coding the think-aloud data

Code	Description
Time	
Collaboration	Time spent on the Cosketch-task, starts from entering Cosketch, and ends when the student starts uploading the chat
Alone in Chat	Time spent in chat, waiting for others
Pre-individual	Time spent on tasks *before* CoSketch-task
Post-individual	Time spent on tasks *after* CoSketch-task
Total	Time spent on the whole (Poem) Human legacy scenario
Non-task	Time spent on activities *not* related to tasks
Pre - individual portion	
2A	Answered questions about the poem students had read
2B	Posed a question about the poem
3A	Copy& paste poem
3B	Score, 1< bubble = 0 point, 1-3 bubbles= 1 p, 3>= 2p
4	Sorted the questions about the poem – more and less relevant
Actions-ColPS	
Roles	
ActualRole_drawer	The student takes on the role as an drawer (only applicable if others are in chat)
ActualRole_observer	The student takes on the role as an observer (e.g. is in the room but do not draw or lead discussions)
ActualRole_writer	The student takes on the role as an writer (e.g. writes about the task suggests how to solve it or help others or clarifies the process, ACTIVE writer role)
ActualRole_sabotage	The student sabotages the drawing or chat by non-task behavior (e.g. messes up the drawing, writes things not related to the task..)
Cognitive skills	
GoalSetting	The student tries to set goals for the task (e.g. explain what needs to be done)
ClarifyProcess	The students tries to clarify either what to draw, the roles (e.g. who will do what) or the progress
TimeManagement	The student asks or posts in the chat time-related information (e.g. time awareness related to finish the task in time)
Draws	The student draws
TaskDiscussion	Communication related to the specific ColPS task
Social skills	
introduceOneSelf	The student introduce him/her when entering the chat (e.g. hi/hello, I am ..)
SeeksToKnowWhosInCh	The students ask whether others are in chat or who is in the chat
SeeksHelpFromOthers	The student ask the others in the chat (e.g. what are we supposed to do, how can I copy the chat,)
GiveSupportToOthers	The student help others (e.g. when someone ask question about what to do or how to do certain things..)
SocialDiscussion	Communication regarding other tasks in the test, or issues related to test situation
Non-task behavior	The student clearly does things **not** related to the test
Incidents and Help	

This is a pre-print version.
DOI to published version: 10.1016/j.chb.2017.08.007

Technical_issues	Every time the student encounters technical issues/problems
Help_tech	Each time the student receives/asks for help to deal with technical issues
Help_task	Each time the student receives/asks for help to deal with task related issues
Post- individual portion	
9A	Searched and found webpage with a poem
9C	Made an audio file
7	Sorted the cards related to the poem
6B	Uploaded the chat from coSketch
9B	Copy and pasted the webpage address (link) to the poem selected
6A	Uploaded the painting from CoSketch
8	Answered questions about the collaboration

This is a pre-print version.
DOI to published version: 10.1016/j.chb.2017.08.007

Table 3

Title?

Id	Gender	Age	Drawing	EvalCollab	FurtherColl	Emoticons	SMS	Text	Grades	Time
Group A										
ST_1	F	14	2	1	F	0	1	2	3,8	12:16
ST_2	M	14	2	1	F	0	2	1	3	17:46
ST_3	M	15	1	1	F	0	1	3	3,8	12:00
ST_4	M	14	2	1	F	0	2	1	3,6	23:13
Group B										
ST_5	M	15	1	2	T	0	1	2	5	41:12
ST_6	M	14	2	2	F	0	0	2	4,3	23:14
ST_7	F	15	1	2	T	0	1	2	5,7	16:19
ST_8	F	14	1	2	T	0	1	3	4,7	28:18
Group C										
ST_9	F	16	2	2	F	0	2	3	4	43:24
ST_10	M	15	2	1	F	0	1	2	3,6	23:51
ST_11	F	16	2	2	F	0	3	3	3,6	24:45

Note. Gender (M= Male, F = Female). Drawing = the sketch, end-product which was uploaded. EvalCollab = evaluation of the collaboration with a score. FurtherColl = Whether the students collaborated after the collaboration task, e.g., for the post-individual tasks. Emoticons = To what extent the students used emoticons for expressing themselves. SMS = To what extent the student used SMS (short message) language. Text = an evaluation of how much text the students wrote in the chat. Grades = average grades of the students on a scale between 2 to 6 (grade 6 is the highest. These were provided by the teachers).

This is a pre-print version.
DOI to published version: 10.1016/j.chb.2017.08.007

Figures

Figure 1. Sample screen from the Human Legacy scenario, page 1

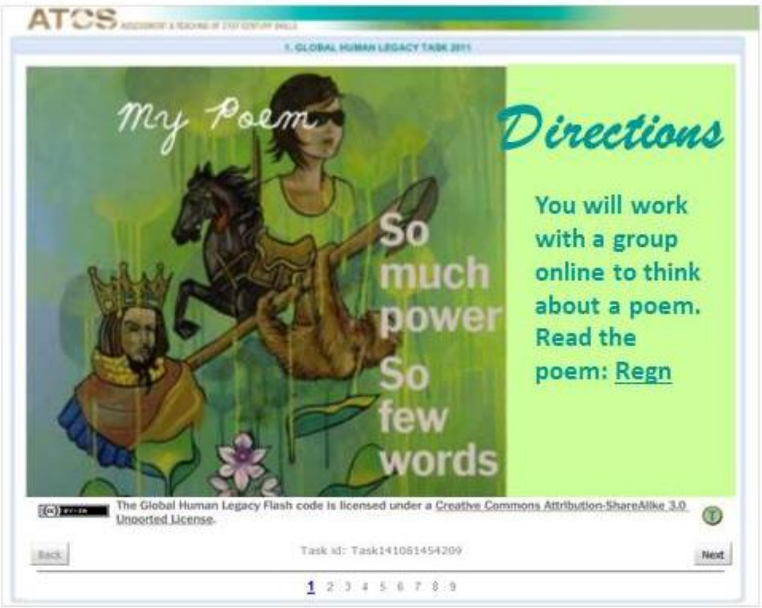


Figure 2

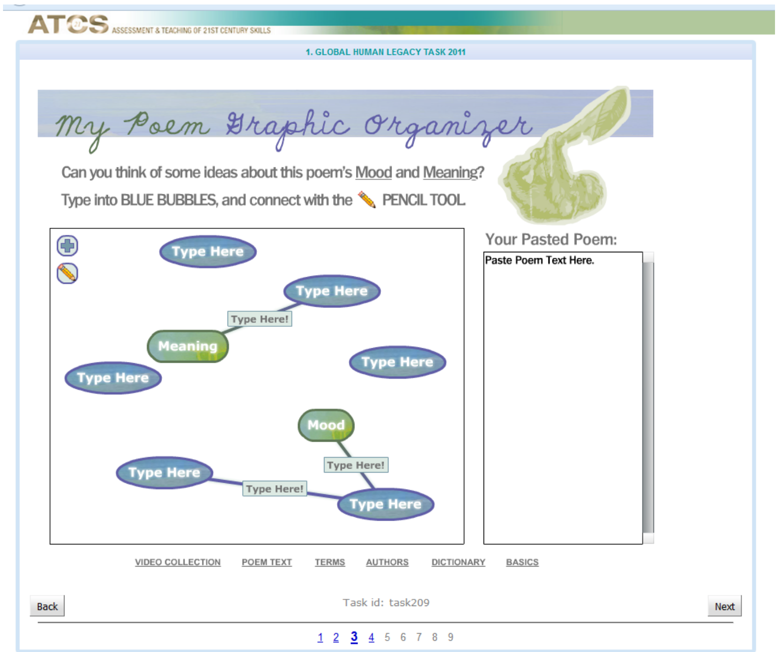


Fig. 2 Sample screen from the mind-map task.

Note. The sample screen shown here is from the English version of the test. The translated version (i.e., Norwegian) was kept the same except the instructions were in Norwegian.

This is a pre-print version.
DOI to published version: 10.1016/j.chb.2017.08.007

Figure 3 Add description

Figure 4. Comparison between group A, B and C.

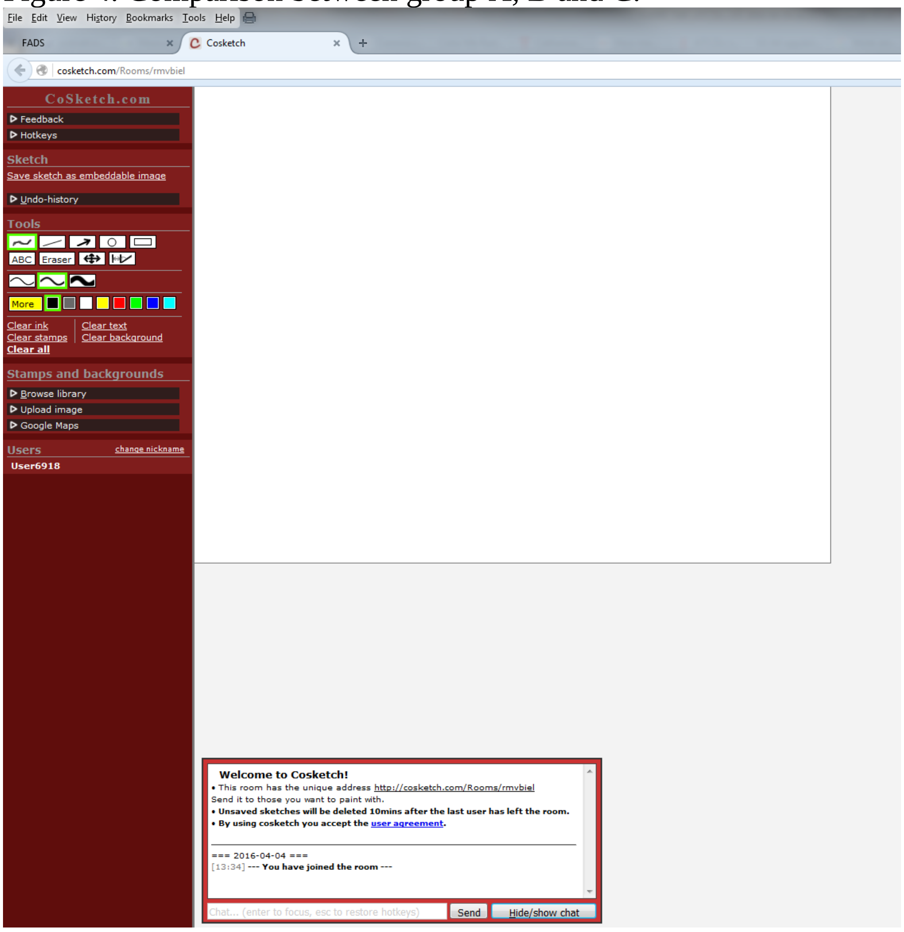


Figure 5.

This is a pre-print version.
DOI to published version: 10.1016/j.chb.2017.08.007

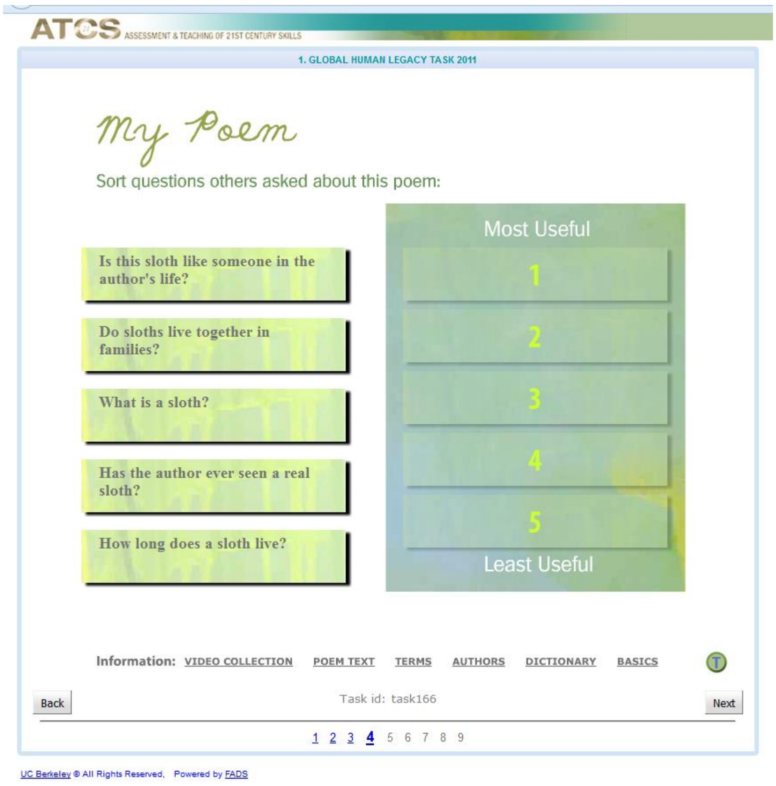
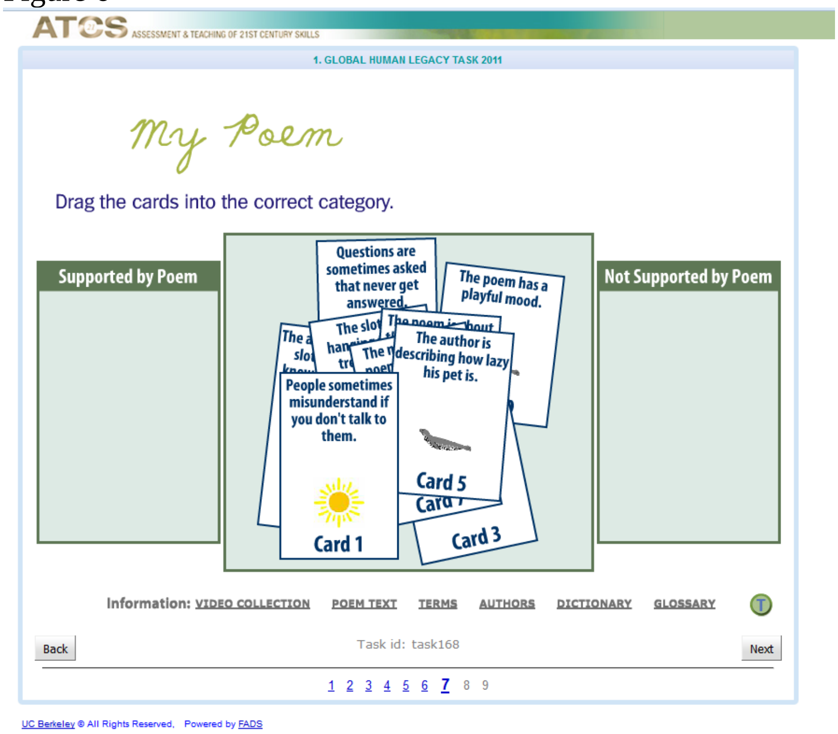
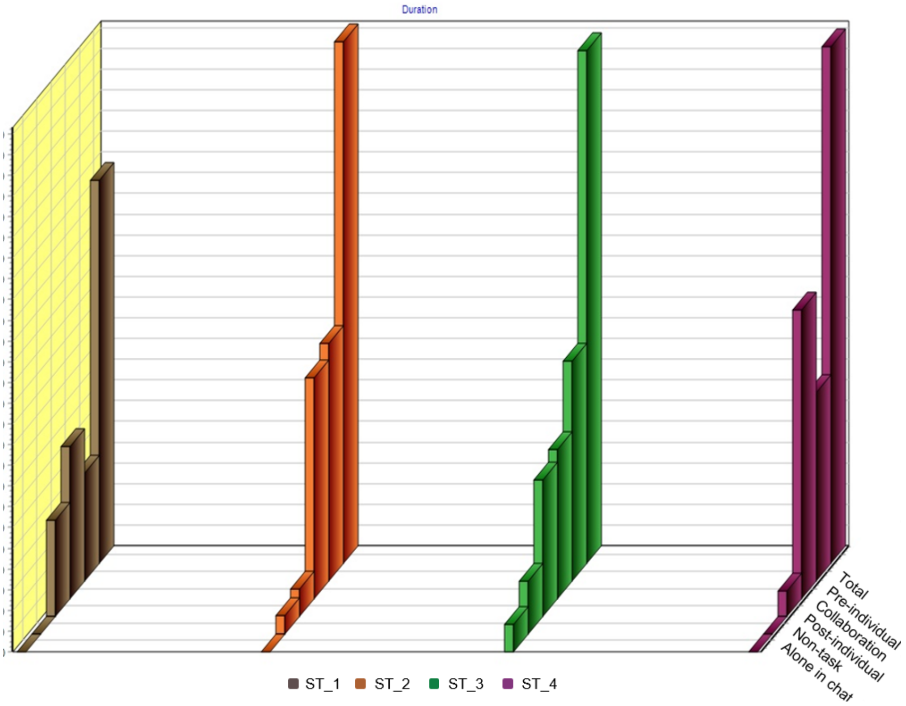


Figure 6



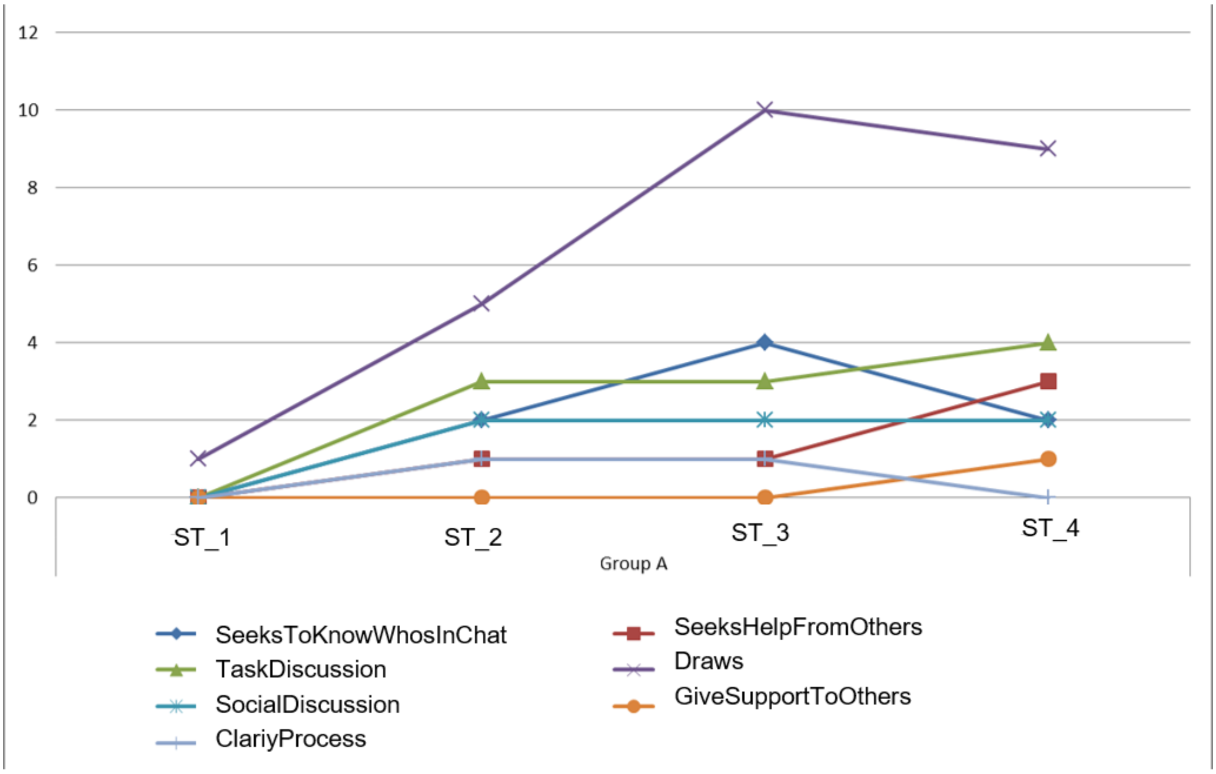
This is a pre-print version.
DOI to published version: 10.1016/j.chb.2017.08.007

Figure 7a An overview of the *Time* category of each student in group A.



The labels "ST_1", "ST_2" and so on represents each student in the group, and each diagram/bar represents the different Time categories as indicated on the right

Figure 7b



DOI to published version: 10.1016/j.chb.2017.08.007

Figure 7c. The collaboration processes of the students in group A

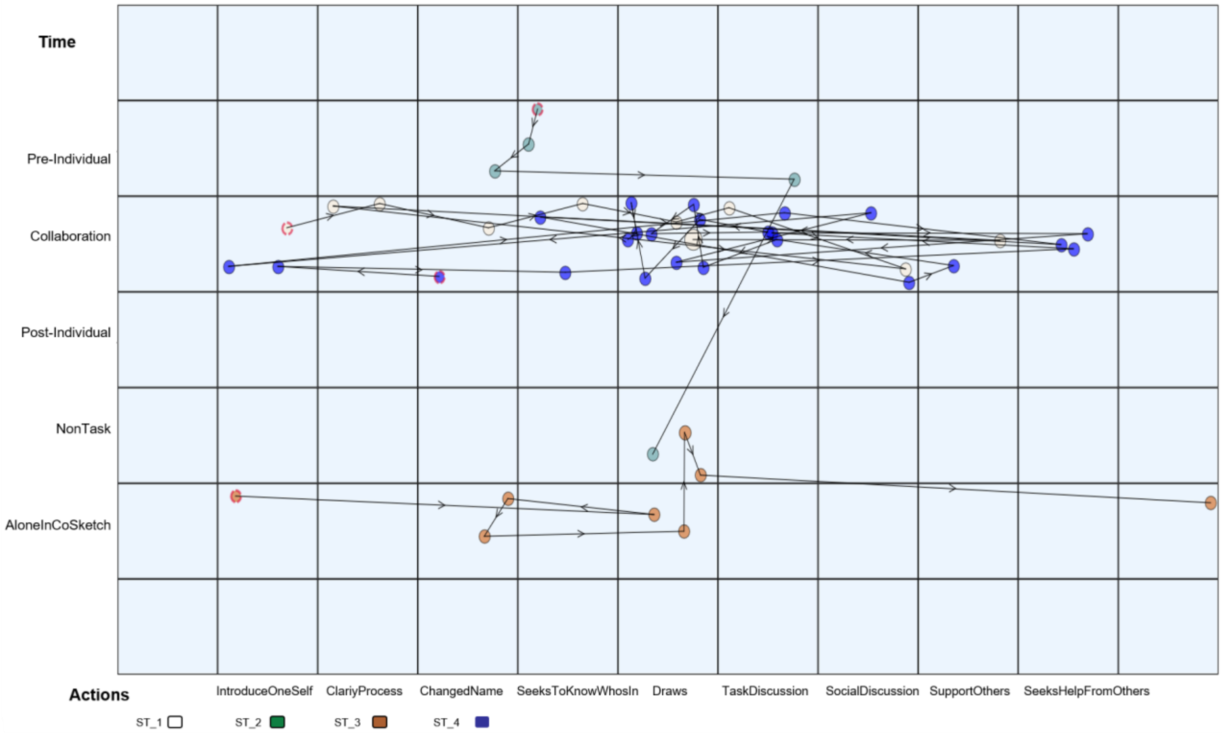


Figure 8a. An overview of the *Time* category of each student in group A

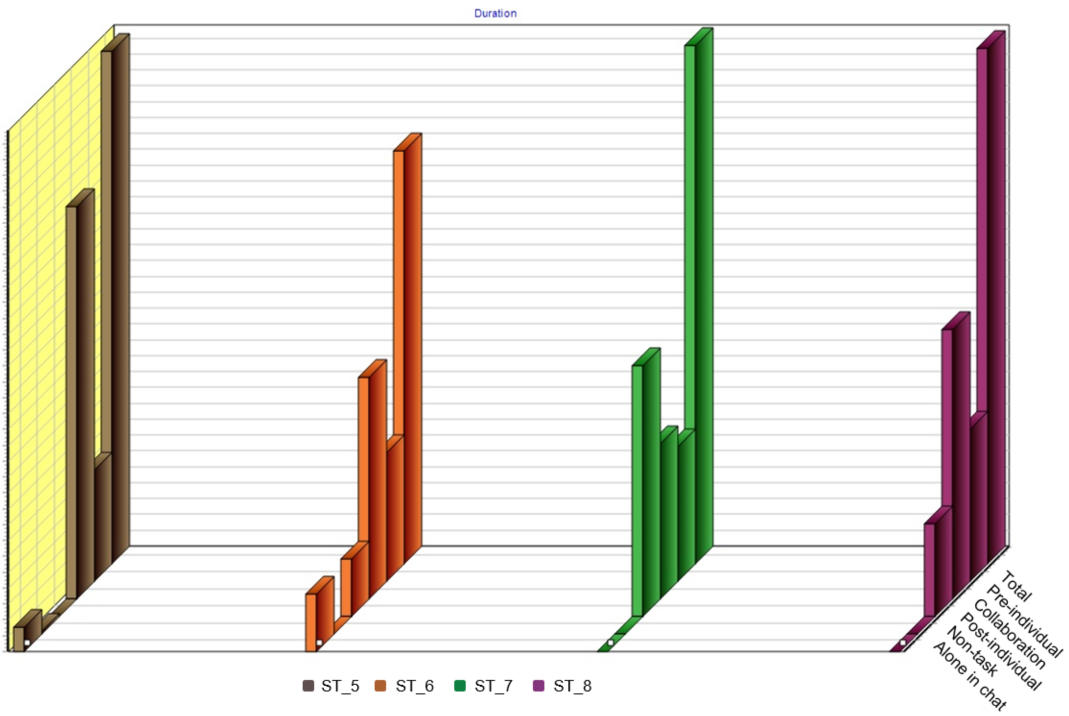


Figure 8b

This is a pre-print version.
DOI to published version: 10.1016/j.chb.2017.08.007

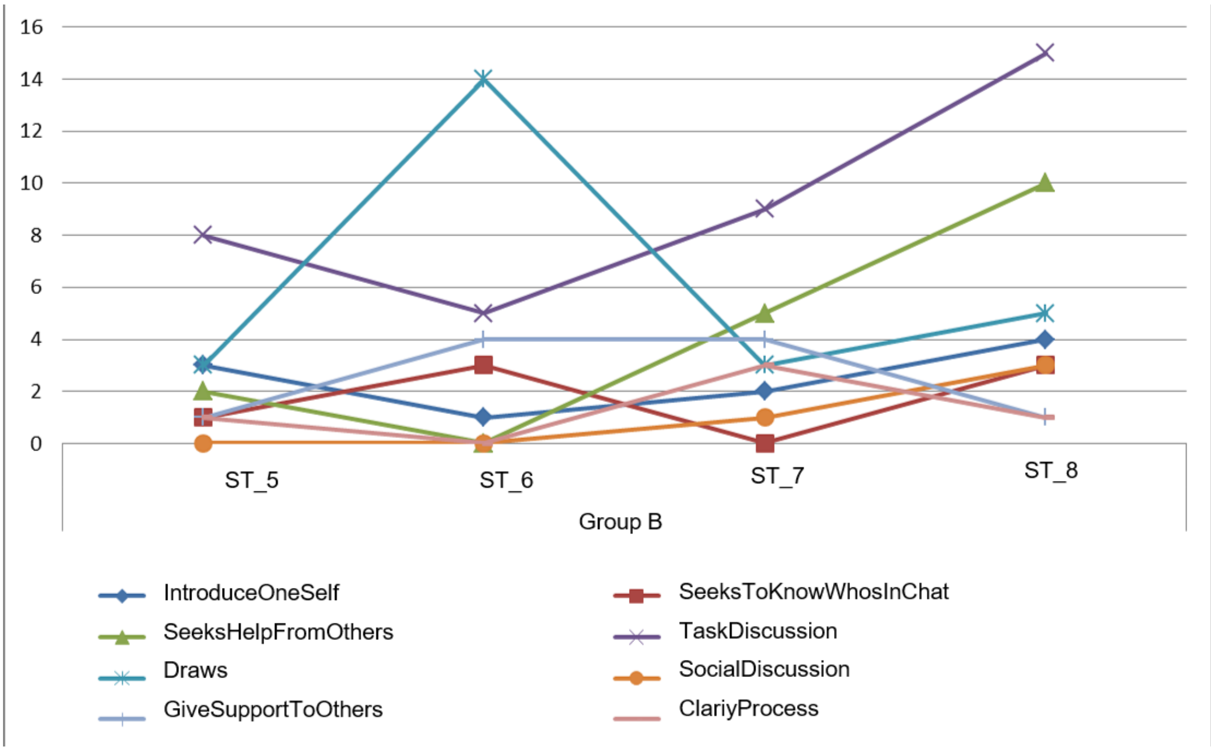


Figure 8c

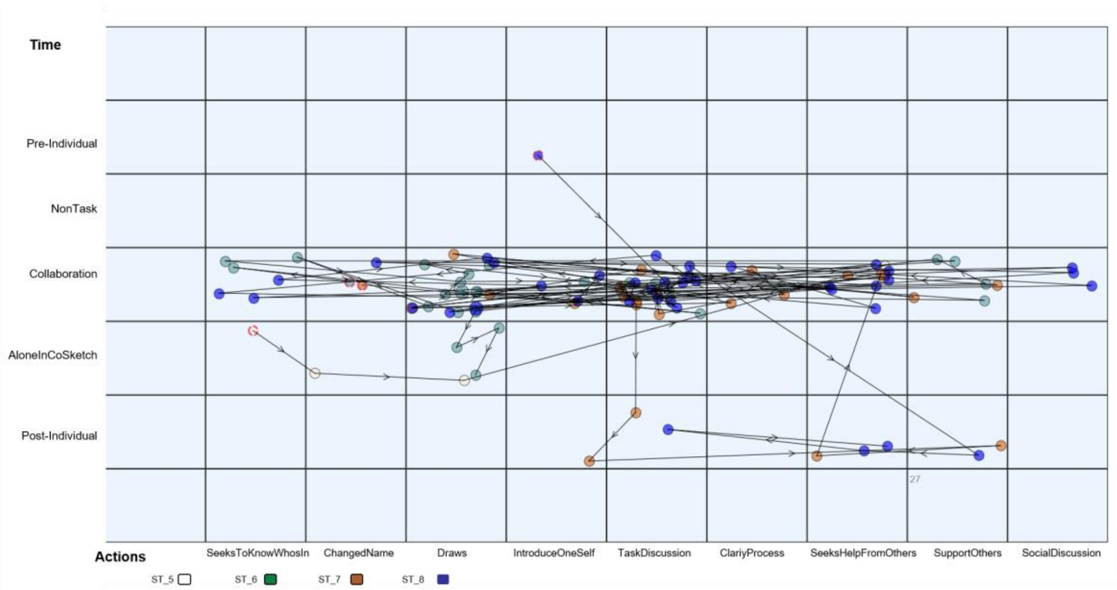


Figure 9a. An overview of the *Time* category of each student in group A

DOI to published version: 10.1016/j.chb.2017.08.007

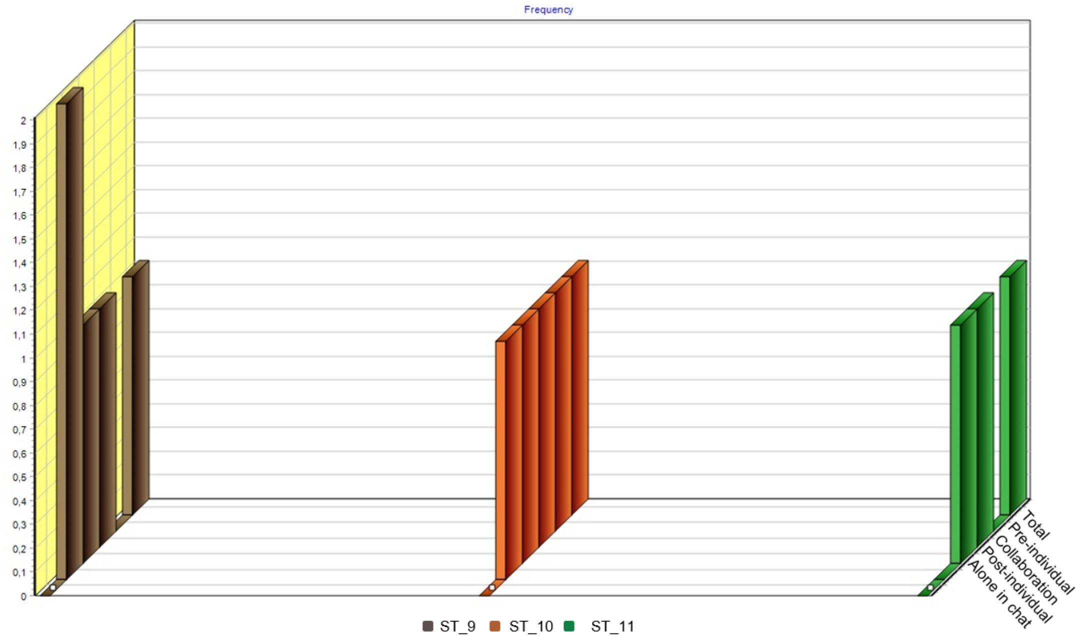


Figure 9b

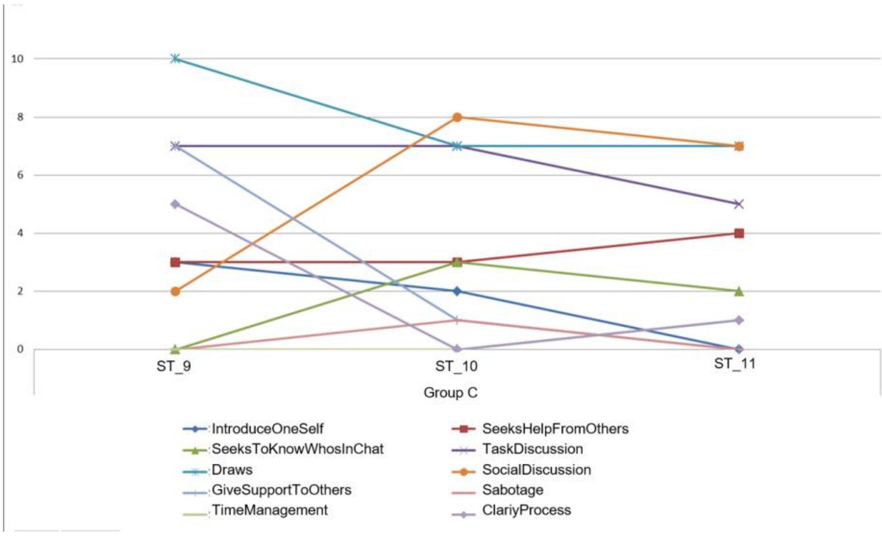


Figure 9c. Y-axis = Time and X-axis= Interaction codes. Student09 =blue; Student10=orange and student11 = green.

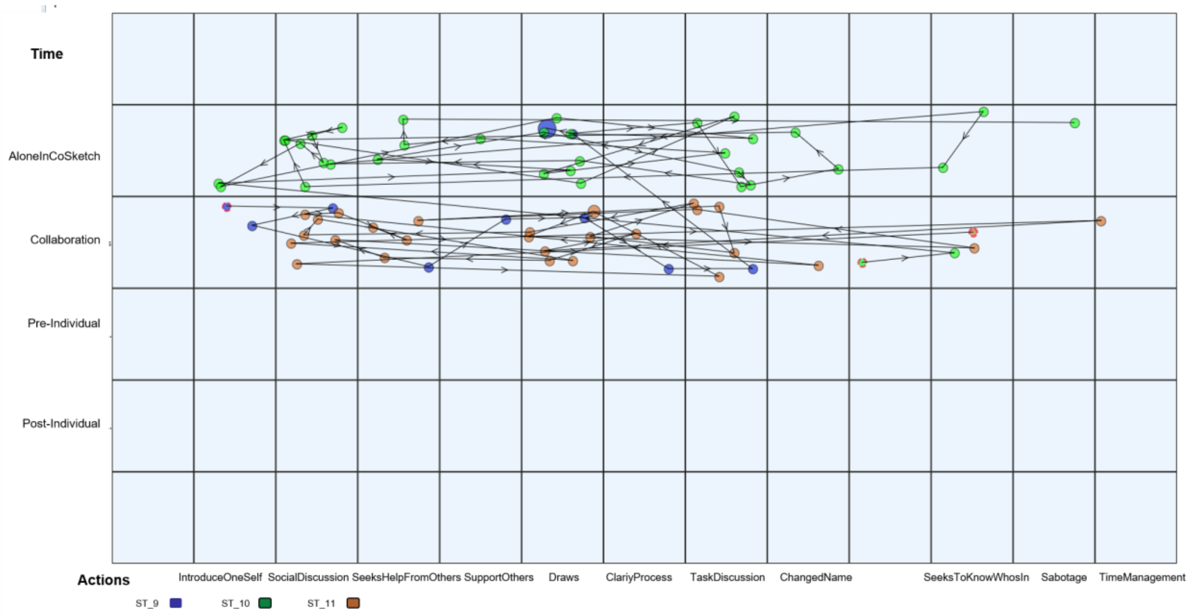
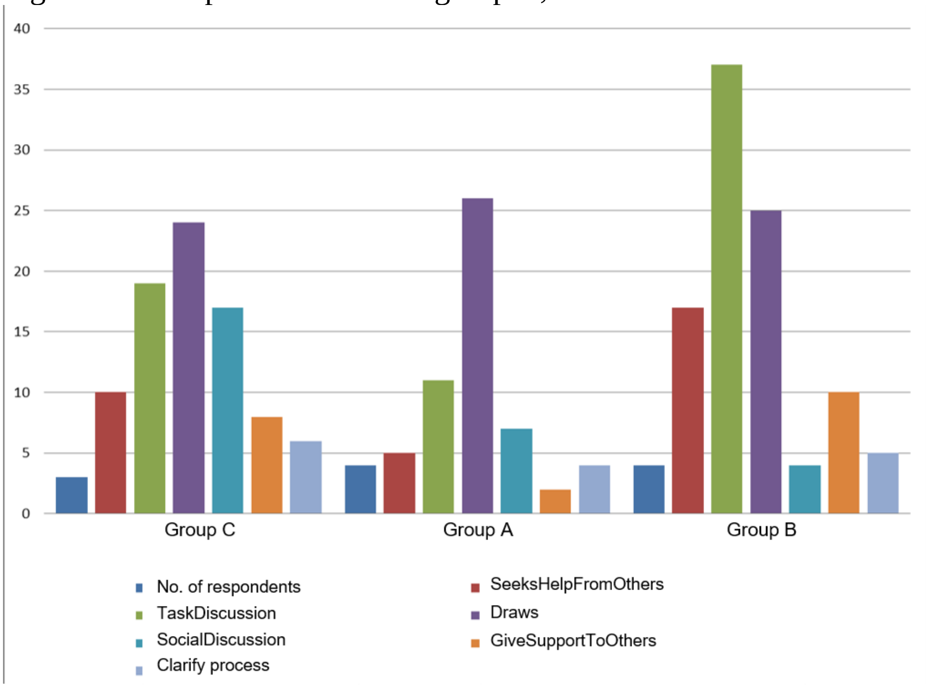


Figure 10. Comparisons between group A, B and C.



DOI to published version: 10.1016/j.chb.2017.08.007

APPENDICES

Appendix 1. Scoring guide

Task id	**Description of task**	**Scoring guide**
2A 2B 3B 4 5 7	Questions about the poem Write down one important question you have about the poem Create a mind map, connect moods and meanings of the poem Sort questions about the poem Drawing/Sketch of the poem (All Group members received the same score – if nothing particular was detected, such as the student did not contribute at all) Drag the cards into correct category	0 = Inaccurate response 1 = Response with low/poor achievement 2 = Response with average achievement 3 = Response with high achievement 99 = Missing/No response given
3A 6A 6B 8 9A 9B 9C evalFurtherColl	Copy and paste the poem text into the box Upload the drawing Upload the chat log Self-evaluation of the collaboration Find a poem you like online. Write the title of the poem. Paste the web address to the poem Upload the audio file you have created The team collaboration further (on other tasks than the ColPS task)	TRUE= Correct response (or Yes) FALSE = Missing, or wrong response (or No)
evalCollab Emoticons SMS-lang Amount_writing	Evalution of the collaboration Use of emoticons SMS* (short message service) language was used How much text the student wrote in chat	0 = Missing, could not be identified 1 = Poor 2 = Average 3 = High

SMS (short message service) is used for text messaging abbreviations or text messaging shortcuts used to reduce typing on your cell phone.

This is a pre-print version.
DOI to published version: 10.1016/j.chb.2017.08.007