Research and Practice in Technology Enhanced Learning
Vol. 9, No. 3 (2014) 367－388

AN APPROACH TO ASSESSMENT OF COLLABORATIVE PROBLEM SOLVING

ESTHER CARE

Assessment Research Centre, University of Melbourne
Parkville Victoria 3010, Australia
e.care@unimelb.edu.au

PATRICK GRIFFIN

Assessment Research Centre, University of Melbourne
Parkville Victoria 3010, Australia
p.griffin@unimelb.edu.au

The Assessment and Teaching of 21st Century Skills (ATC21S™) project had a research and development plan that reflected a traditional approach to conceptualisation of constructs, and then establishment of test blueprints followed by test construction. The test construction process included several stages – of drafting task ideas, panelling these with teachers, then piloting revised tasks, and trialling these with students. Within the conceptualisation phase, the project focused on the definition of 21st century skills. This paper outlines the process of moving from the definition to the assessment of one such set of skills – collaborative problem solving. It describes how this led to the development of hypothesised learning progressions which portray how the skills might vary across more and less adept individuals. In the process of development of collaborative problem solving assessments, the tasks were reviewed in terms of how they might reflect the construct and how they might engage students such that their thinking processes could be captured, coded, scored and interpreted. The overall method of this process together with its rationale is described.

Keywords: Human-to-human; 21st century skills; problem solving.

1. Introduction

Structural changes in employment in many developed countries support the need for system-wide change in education. Some skills needed in the 21st century are not included explicitly in many curricula, although there are many countries moving in that direction. Australia and Singapore for example, have outlined the skills that these countries believe important for development in their students. Unlike these systems' specifications of how traditional disciplines might be taught and assessed, the guidelines provided for the teaching of Australia's "general capabilities" and for Singapore's "21st century competencies" are general, and for assessment are comparatively lacking.

Assessment approaches and formats traditionally associated with education may not be capable of capturing performances of 21st century skills, especially those that might be considered non-cognitive. By traditional, we refer to assessments that value correct answers or answers that conform with a view about "ideal" approaches to a question or issue. Assessment types in education frequently use closed questions or items where a response is considered either correct or incorrect. Focus on affective characteristics, or on inferences about skills levels to be drawn from performance, which might lead away from the correct/incorrect dichotomy, is less common. Mindful of this consideration, new assessment approaches commensurate with the nature of 21st century skills (Griffin, McGaw, & Care, 2012) need to be considered. The context for this approach is a concern not only with how best to assess such skills but how to describe them sufficiently for teachers to develop in their students.

Many skills have been identified by many authorities as 21st century skills (Table 1). These perspectives became evident in the 20th century, as can be seen by three initiatives undertaken by the United Nations Educational, Scientific and Cultural Organization (UNESCO), the Organization for Economic Cooperation and Development (OECD), and the European Commission. These provide a useful context for the practical work on teaching curriculum and assessment in 21st century skills. UNESCO took a competence approach. The Delors' Report (1996) marked the beginning of UNESCO's 21st century competence learning discourse - with learning to know, learning to do, learning to be, and learning to live together - forming the four pillars of learning. Learning to know includes developing the faculties of memory, reasoning and problem solving; it pre-supposes learning to learn and could usefully be extended to the concept of knowledge building. This perspective does not presume that knowledge is fixed. Learning to do implies acquisition of complex skills, but also refers to developing an aptitude for teamwork and initiative, and a readiness to take risks. Learning to live together is the pillar UNESCO emphasizes more than any other. It refers to developing an understanding of others as well as highlighting the reality that if we are to understand others, we must first know ourselves. Learning to be is founded on the fundamental principle that education needs to contribute to the all-round development of each individual. This pillar deals with what it is to be human, comprehended by intellectual, moral, cultural and physical dimensions.

The OECD's position, developed within the DeSeCo Project - Definition and Selection of Competencies (Rychen & Salganik, 2003; OECD, 2013) - has a focus on key competencies and classifying these competencies in three broad categories. First, individuals need to be able to use a wide range of tools for interacting effectively with others and the environment. They need both physical tools such as information technology and socio-cultural ones such as the use of language. They need to understand these tools well enough to adapt them for their own purposes. Second, in an increasingly networked and interdependent world, individuals need to be able to engage with others. Third, individuals need to take responsibility for managing their own lives through situating themselves in the broader social context.

The European Commission's paper (Gordon et al., 2009) highlights the cross-curricular relevance of key competences such as mathematics, digital competences, individual and social responsibilities and cultural awareness. Gordon et al. draw out issues about key competences both for students and their teachers, as well as alignment across the different education sectors and employment. Beyond the cross-curricular areas, the importance of communication both in mother and foreign tongues is reviewed. This report moves beyond the cataloging of competences to an in-depth treatment of implications for teaching and policy.

Partnerships 21 took as their mission to catalyse the United States' (US) Kindergarten to Year 12 (K12) education for the 21st century. Essentially they endorse the "fusing" of traditional academic disciplines (reading, writing and arithmetic – the 3Rs) with skills including critical thinking, communication, creativity, and collaboration - the 4Cs. These are contextualised within life and career skills, and technology and media skills.

The concern of ATC21S was not only with the definition and identification of 21st century skills (Binkley et al., 2012), but with the methods appropriate for assessment of these (Wilson, Bejar, Scalise, Templin, Wiliam, & Torres-Irribara, 2012), the types of technologies on which these might depend (Csapó, Ainley, Bennett, Latour, & Law, 2012), the teaching approaches that might be deployed (Scardamalia, Bransford, Kozma, & Quellmalz, 2012), and the implications of these for policy change (Darling-Hammond, 2012). From the broad-ranging and comprehensive review of these issues, ATC21S then turned to demonstrating the implications of these theoretical perspectives through its focus on collaborative problem solving (Griffin & Care, 2015).

These five perspectives should not be under-estimated in terms of their influence on government policies worldwide, and on the education sector. Although the definitions and descriptions of concepts discussed at policy level may well end up the object of dissection in academic literature, as well as verification through academic research, more often than not the actual field of practice moves ahead of such activities. This is a conundrum in scientific method – that the world cannot wait for the research to inform movement, and instead relies on what appears to be well-reasoned and logically appealing arguments.

Each of the approaches (Table 1) to understanding of 21st century skills and how they fit with our notions of education and the function it serves, emphasises skills that have not been explicit in descriptions of traditional education disciplines. They all actually identify enabling skills - skills that we need to navigate our global society. They converge on a common set of 21st century competencies - collaboration, communication, Information Communications Technologies (ICT) literacy, and social and/or cultural competencies; and most include creativity, critical thinking, productivity, and problem solving.

The competencies have individually been of great interest across different environments. For example, creativity has been the object of much academic research; collaboration has been of interest in industry, in the military, and in human relations environments. The competencies have not, however, been highlighted until recently in the education sector – it having typically focussed on cognitive skills and achievement connected with traditional literacies such as language, mathematics, science, and the social sciences.

Table 1. Comparison of classifications of 21st century skills.

ATC21S	UNESCO	OECD	Partnerships 21	European Commission
Binkley et al. (2012)	Delors et al. (1996)	OECD (2013)	www.p21.org	Gordon et al. (2009)
Ways of thinking	**Learning to know**		**Learning and innovation**	**Learning to learn**
creativity and innovation critical thinking, problem solving, decision making learning to learn, metacognition			creativity critical thinking problem solving	
Ways of working	**Learning to do**	**Interact in heterogeneous groups**		
communication collaboration		relate well to others co-operate, work in teams manage and resolve conflicts	communication collaboration	communication in mother tongue and foreign languages
Tools for working	**Learning to do**	**Use tools interactively**	**Information media and technology**	
information literacy ICT literacy		use language, symbols and texts interactively use knowledge and information interactively use technology interactively	information literacy media literacy ICT literacy	mathematical, science and technology competences digital competence
Living in the world	**Learning to be** **Learning to live together**	**Act autonomously**	**Life and career**	
citizenship - local and global life and career personal and social responsibility - including cultural awareness and competence		act within the big picture form and conduct life plans and personal projects defend and assert rights, interests, limits and needs	flexibility and adaptability initiative and self-direction social and cross-cultural skills productivity and accountability leadership and responsibility	social and civic competences initiative and entrepreneurship cultural awareness and expression

Note. Skills in bold represent categories of skills within the specified framework

2. Collaborative Problem Solving

Collaborative problem solving is one skillset that has been identified as of interest as a 21st century skill. It is a skillset that has not sat clearly in one particular field of research. It has been of interest in health, psychiatry and social welfare contexts (e.g. Pollastri, Epstein, Heath, & Ablon, 2013); human resources and team building contexts; and most recently has emerged in the Assessment and Teaching of 21st Century Skills project (Griffin et al., 2012), and subsequently in the large scale testing area (e.g. Greiff, 2013) highlighted by OECD's Programme for International Student Assessment (PISA) initiative. Both these initiatives have moved beyond definition and description to development of assessment methods and tasks.

Collaborative problem solving is conceptualised as a complex skill requiring both social and cognitive competencies. It arises from the links between critical thinking, problem solving, decision making and collaboration – all of which processes are hypothesised to contribute to the skill. O'Neil, Chung, and Chuang's (2003) seminal work in exploring collaborative problem solving in computer-based assessment environments included the conceptualisation of collaborative learning, problem solving, and higher order capacities. Earlier, O'Neil (1999) defined problem solving as requiring content understanding, problem solving strategies, and regulation of one's problem solving processes and progress.

Hesse, Care, Sassenberg, Buder, and Griffin (2015) conceptualise collaborative problem solving as structured across five broad strands – participation, perspective taking, social and task regulation, and knowledge building – set within two organising components, cognitive and social. They define collaborative problem solving as a set of skills on which individuals need to rely when the capacities or resources of just one person are not sufficient to solve a problem. The skill lies in how to combine different resources and skills when faced with complex problems.

Hesse et al.'s (2015) framework echoes components of O'Neil and colleagues' description of collaborative problem solving, but differs from this in terms of more clearly defined social or collaborative skills, and in terms of identifying the relevance of regulation to both the social and cognitive dimensions. Social regulation includes exerting skills to manage the interpersonal space that is initiated by the problem solving interaction, while task regulation refers to the skills required to map out the problem space itself – the questions that it poses, the resources or artefacts within it, and the processes that might be followed. The framework is conceptualised at three levels with collaborative problem solving as the over-arching construct, to which two components – the social and cognitive – contribute. Within each of these components, are subskills, termed elements. In Table 2, these elements are expanded upon in the Indicator column to make clear what actions or processes they comprehend. It is these actions or processes that assessment tasks must be constructed to require in order for the measurement to reflect the construct.

Table 2. Collaborative problem solving framework.

Strand	Element	Indicator
Participation		
	Action	Acts within environment
	Interaction	Interacts with, prompts and responds to the contributions of others
	Task completion / perseverance	Undertakes and completes a task or part of a task individually
Perspective Taking		
	Adaptive responsiveness	Ignores, accepts or adapts contributions of others
	Audience awareness	Awareness of how to adapt behaviour to increase suitability for others
Social Regulation		
	Negotiation	Achieves a resolution or reaches compromise
	Self evaluation	Recognises own strengths and weaknesses
	Transactive memory	Recognises strengths and weaknesses of others
	Responsibility initiative	Assumes responsibility for ensuring parts of task are completed by the group
Task Regulation		
	Problem analysis	Analyses and describes a problem in familiar language
	Sets goals	Sets a clear goal for a task
	Resource management	Manages resources or people to complete a task
	Flexibility and ambiguity	Accepts ambiguous situations
	Collects elements of information	Explores and understands elements of the task
	Systematicity	Implements possible solutions to a problem and monitors progress
Learning and Knowledge Building		
	Relationships (Represents and formulates)	Identifies connections and patterns between and among elements of knowledge
	Rules "If … then"	Uses understanding of cause and effect to develop a plan
	Hypothesis "what if …" (Reflects and monitors)	Adapts reasoning or course of action as information or circumstances change

Note. Adapted from Hesse, Care, Buder, Sassenberg, & Griffin (2015)

3. Framework for Task Construction

Collaborative problem solving means approaching a problem proactively and responsively, by working together and exchanging ideas. Collaboration is a useful tool, especially when specific expertise is needed (and available), and relies on factors such as a readiness to participate, mutual understanding, and the ability to manage interpersonal conflicts. Collaborative problem solving is particularly relevant when dealing with a complex problem.

Collaboration in the context of problem solving involves the search for relevant information from another person, joint use of different resources and agreement on strategies and solutions. It requires the active participation and responding with other people as well as taking others' perspectives and evaluating self and peers in the context of capacity to contribute.

Social and cognitive components contribute to the hypothesised structure of collaborative problem solving. These two components are not totally independent of each other, and the degree to which their interaction in a collaborative problem solving environment modifies both social and cognitive functioning is not yet known. Notwithstanding, collaborative problem solving is seen as not just an extension of individual problem solving but as a construct in its own right. One of the primary differences between individual and collaborative problem solving is that the latter must be explicit or visible given the major need for communication and sharing of information. The degree to which this visibility actually modifies the process of reasoning remains to be ascertained. In the Hesse et al. (2015) framework, the social component draws on literature from social and organisational psychology while the cognitive component draws heavily on classical approaches to individual problem solving. Simply stated, social skills are about managing the collaborating individuals, and the cognitive skills are about managing the task.

3.1. *The social component*

Collaboration is the activity of working together towards a common goal. There are a number of activities included in the definition. One activity is communication, the exchange of knowledge or opinions to optimise understanding by a recipient. This is a necessary but not sufficient condition for collaborative problem solving – communication must go beyond mere exchange. Individuals need to be able to take the perspective of others, and provide responsive contributions. Another activity involves managing the collaboration itself – the skills of working with others and participating. This conceptualisation of the relevant social skills refers to three classes of indicators - participation, perspective taking, and social regulation.

Participation refers to an individual's readiness to externalise and share information and thoughts, and their actual involvement. Perspective taking skills enable an individual both to understand another's point of view, and to modify or adapt their own behaviours in light of this recognition (Dehler, Bodemer, Buder, & Hesse, 2011). Social regulation skills provide the facility for individuals to be aware of and manage the problem space in terms of the implications of human behaviour upon it. Intra and inter-personal awareness is essential for optimising these strategic aspects of collaborative problem solving (Peterson & Behfar, 2005).

3.2. *The cognitive component*

The problem solving component to a large degree relies on reasoning skills, as identified through an information processing approach. Both the strands of Task Regulation and

Knowledge Building rely on these skills and the presumption is that reasoning can be taught. In this context it becomes an important component in problem solving, and more broadly in collaborative problem solving. Reasoning can be seen as occurring across inductive and deductive processes. Inductive reasoning focuses on finding patterns. This relies on exploration of information and identifying connections between artefacts in the problem space (Fischer, Greiff, & Funke, 2012). Deductive reasoning focuses on understanding the implication of logic statements and rules. According to Polya (1973), problem solving can be seen as a sequence of processes – such as understand the problem, devise a plan, carry out the plan, look back and check. The presumption that these skills can be taught is informed by the degree to which empirical evidence of increasing sophistication of the skills occurs, and is an important aspect of task design.

While inductive reasoning focuses on establishing a possible explanation to test in the first place, deductive reasoning involves testing whether the explanation is valid or not. This combination, described by Griffin (2014), is known as hypothetico-deductive reasoning and is often taken to be the defining characteristic of scientific practice. Intrinsic to this process is the role of theory. The first step in this method is to generate hypotheses based on the theory. The next step is to perform experiments (or take systematic and controlled observations) that permit testing of these hypotheses. The deductive method attempts to "deduce" facts by eliminating all possible outcomes that do not fit the facts.

In order to measure this set of subskills or processes comprehended by the collaborative problem solving construct, an approach to assessment, and the method followed to construct a series of online tasks is outlined. The online tasks are created to enable pairs of students to engage in collaborative problem solving activities, and to capture the thinking processes used by the students which act as indicators of the strategies used.

4. Method

4.1. *Approach to development of assessment tools*

Wilson (2005) argued that hypothesising of a developmental progression that underlies a construct is a necessary step in constructing assessments. In order to do this the construct of interest must be described with detail about any components or subcomponents. The progression itself can be used by task developers to identify the type of content required to sample the construct. In order to construct a hypothetical progression, Wilson suggests outlining at least three levels of performance that will illustrate the difference between those who possess a great deal of the construct and those who possess little. In this way, the order, direction and magnitude of skill possessed by individuals can be mapped out. The task developer can then draft ideas about how the skills might be demonstrated, and then design a task to elicit these in a way that they can be described along the hypothesised progression.

The development of assessment tools starts with the imperatives that the hypothesised constructs are theoretically sound; that the evidence of growth in skills can be mapped onto a developmental progression; and that a framework can be established within which the assessment tasks and consequent teaching strategies can be conceptually based.

In this case of exploring collaborative problem solving and developing tasks to measure it, its contextualisation as a 21st century skill and a skill that would lend itself to teaching and learning in the classroom, was important. The steps of conceptualisation of collaborative problem solving, its description, hypothesised progression, and development of tasks to measure it, were undertaken with three main purposes in mind. First, the process itself should serve to clarify the nature of the construct in such a way that demonstrations of the skillset would be recognisable; second, the process should lead to a prototype set of tasks that would act as a model for implementation in an online environment to take advantage of digital technologies to track students' thinking processes; and third, the process should provide teachers with an approach for its teaching.

In terms of product, assessment tasks need to provide evidence to bring to the hypothesised developmental progression. Since in this case, the construct of interest is a general skill, its measurement needs to be distinguished from the context in which tasks are set. To achieve this, tasks need to be developed that embed the skills in different content areas, and in areas that will vary in terms of how much content knowledge is necessary. Individuals' capacity to exercise their skills across content and context would demonstrate the degree to which the skills are generalisable. It might not be intuitive to relate novice levels of skills developed in the classroom, to the expert level of skills that one would associate with complex problems in the real world. Therefore, the definition of the construct itself must ensure congruence of the construct with the assessed reality. The linking of construct identification and assessment approach is intrinsic to this work since the assessment itself is a validation tool for the definition and description of the construct. The classroom provides an environment amenable to collaborative work. The challenge for the design of online assessment tasks is to transfer the physical observation and measurement of participation in a task of a class of students working together, discussing the problem, and talking to each other, to a technology context – with data on the activity captured in the background. The challenge is to identify and record those procedures so that they can be coded and interpreted in ways that are useful for teaching and learning.

4.2. *Functions and characteristics of the tasks*

Eleven human to human collaborative problem solving tasks were designed. The intent of the process was to identify ways to assess these skills and in so doing provide a tool for teachers to enhance their understanding of the construct and to facilitate its teaching. For these latter purposes, there was also preparation of professional development modules, and a series of student and school reporting modules. Each of the tasks was developed independently of each other, without a predefined template. All 11 tasks were taken through the process described below.

There was a requirement that the assessments have the capacity to monitor student differences across lower to higher-order performance. The skills underpinning the assessment tasks are expected to be teachable and measurable. The nature of problems that require collaborative problem solving skills in real life is typically ambiguous, requires multiple resources (skills, knowledge, artefacts) and the engagement of individuals who are dependent on one another for successful resolution. Further, some tasks should be asymmetrical, with each of the students accessing or controlling different but essential data and resources, reflecting the nature of the problems for which the construct is deemed relevant.

To illustrate the design process, the Laughing Clowns task is described. In its most simple iteration, this task would require a problem solver to observe a circus or carnival game stall in which an automated clown head moves from side to side as an individual throws balls into its mouth. The task for the problem solver is to identify if there is a consistent pattern which describes where a ball lands as it comes out of the bottom of the clown's head according to its position at the point when the ball was thrown into its mouth. The positions at which the ball could enter the clown's mouth might be at the left, the centre, and the right. The components of the task are then identifying if entry at each of these points results in the ball exiting systematically at one of three (or more) positions – for simplicity's sake, at position 1, 2 or 3. In summary the processes are:

- collect information: drop balls in each position, note position; observe exit point, note position
- identify patterns: do balls come out in a predictable way?
- form rules: left -> 1, centre -> 2, right -> 3
- test rule: e.g. is it always L-1?

The task was adapted for the assessment of collaborative problem solving skills. Beyond collecting information, identifying patterns, forming and testing rules, the task can be extended to measure rule generalisation, and generation and testing of hypotheses. Beyond these cognitive skills, the creation of two instead of one clown machines and two instead of one problem solver, requires collaborative skills to be brought into play.

The Laughing Clowns task involves students finding patterns, sharing resources, forming rules and reaching conclusions. Two students are each presented with a clown machine and 12 balls to be shared between them. The clown heads move back and forth; when a ball is placed in the mouth, it comes out through one of three shutes at the bottom of the clown head. The task is to determine whether the two clown machines work in the same way. The students need to share information and discuss the rules as well as negotiate how many balls they should each use. The students must place the balls into their clown's mouth while it is moving in order to determine the rule governing out of which shute the balls will emerge (positions at the Left, Middle, and Right – L, M, R). Each student must then indicate whether or not they believe the two machines work in the same way. This adaptation of the simple task requires the collaborative skills of participation, perspective taking, and social regulation.

In developing tasks such as the Laughing Clowns, test developers mapped the problem solving and collaborative elements from the Hesse et al. (2015) framework to the processes individuals would need to undertake to complete the task. Figure 1 illustrates the development process. The first step was to suggest scenarios which might sample the skills described in the framework, and capture these across the hypothetical progressions. The context for each task needed to be identified in order to create artefacts which would act as resources and prompts in the problem space posed by the task itself. This problem space needed to be able to generate sufficient data to enable the interpretation of student performance. The performance needed to reflect the collaborative problem solving construct - the major social and cognitive skills dimensions, and the five strands (Participation, Perspective Taking, Social Regulation, Task Regulation, and Learning and Knowledge Building).

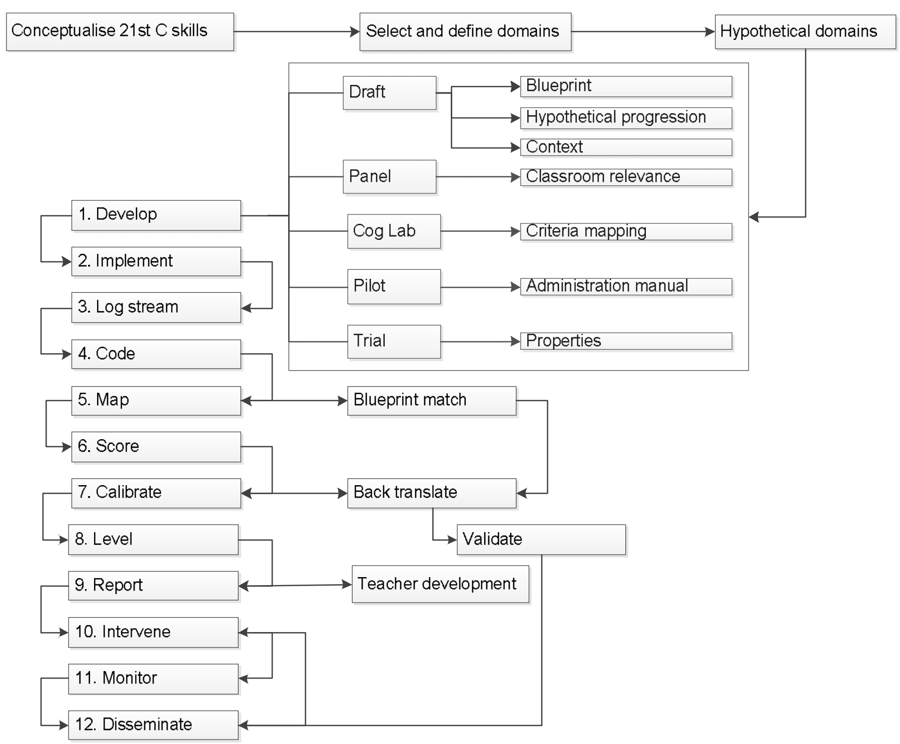

Figure 1. The process.

Table 3. Section of the blueprint for the Task Regulation strand.

Element	Indicator	Low level	Middle level	High level
Organises (problem analysis)	Analyses and describes a problem in familiar language	Problem is stated as presented	Problem is divided into subtasks	Identifies necessary sequence of subtasks
Flexibility and ambiguity	Accepts ambiguous situations	Inaction in ambiguous situations	Notes ambiguity and suggests options	Explores options
Systematicity	Implements possible solutions to a problem and monitors progress	Trial and error actions	Purposeful sequence of actions	Systematically exhausts possible solutions

Note. Adapted from Hesse, Care, Buder, Sassenberg, & Griffin (2015)

The blueprints provided guidance to test developers in terms of the types of behaviours hypothesised to indicate the subskills, and to demonstrate different levels of quality or performance. Table 3 provides an example of nutshell descriptions of how students at different levels of the three elements of "Organises", "Flexibility and Ambiguity" and "Systematicity" within the Task Regulation strand might approach a task. The drafting of tasks needed to take these possibilities into account. Once the tasks were drafted, they were taken to research participants for panelling, piloting and trialling. These steps in task development were interwoven with coding, scoring and calibration activities.

4.3. *Participants*

Students, teachers and the schools which contributed to this research were recruited through the Assessment and Teaching of 21st Century Skills (ATC21S^{TM}) project. All students and teachers comprised convenience samples. Each participating country recruited their participants according to their own academic research and/or governmental research protocols and ethics conventions. The students contributing data to the analyses presented in this article, numbered 4056 adolescents from Australia, Costa Rica, USA, Finland, the Netherlands, and Singapore. Students were aged in the main between 13-17 years of age. The teachers contributing to the steps were the teachers of these students. The first step of panelling, or checking by the teachers, was undertaken on English versions of the tasks. Subsequent steps with students were undertaken on English, Spanish, Finnish and Dutch versions of tasks.

4.3.1. *Drafting and checking*

Draft tasks were taken to teachers to check content and curriculum relevance, usability, and whether the tasks would differentiate between students who are less and more skilled. Teacher advice was sought about likely student engagement and how formative feedback might be provided that would support teaching and learning. The consequent information was fed back to task developers regarding the creation of a suitable interface for teachers and students to interpret and use for the promotion of developmental learning. Much of the information derived from the teachers centred on the opportunities provided by the tasks for students' reasoning to be clear. The other major feedback revolved around the lack of clarity of the tasks in terms of specification about how to engage with them and work through them.

4.3.2. *Cognitive laboratories*

Cognitive laboratory materials were designed to enable the collection of information primarily about how students approached and worked with the task. Of interest were the ways in which students thought through and interacted during the problem solving process. The information enabled task developers to review the kinds of coding that might be used to record the students' actions. For participation, students were identified by the teacher from the lower, middle and upper sections of the class based on general academic abilities. This was done to optimise the chances of exploring how students across ability ranges might engage with the tasks, and how mixed ability pairing might affect estimates of student ability. Much of the feedback from students obtained in focus groups after the cognitive laboratories centred around the lack of scaffolding of their experience, and how they were challenged to make sense of an ambiguous environment. Both teacher and student feedback to this point reflected expected responses to the nature of the collaborative problem solving task environment which was deliberately constructed to mirror the nature of the construct itself.

4.3.3. *Pilot*

To ensure that the tasks could be administered and undertaken at full class levels, the pilots were implemented in the participating countries. This required the hosting of the interactive tasks on a browser-accessible platform. Of interest was the degree to which classroom administration could be managed by the teacher, required time for completion of tasks, and the capacity of the system to manage the online synchronising of information as all students were working. Accordingly, the tasks were presented to classes of students in a format similar to that envisaged for the final tasks.

4.3.4. *Trial*

The trials were implemented across the six countries, with online registration facilities inbuilt to the task environment, which latter had the capacity to capture the online collaborative interaction between students. The data captured from the 4056 students were used to develop empirically-based scales that would have the capacity to locate students' on the learning progression. This process established the psychometric properties of the tasks.

4.3.5. *Data capture*

The assessment tasks used data capturing processes that allow the recording of all actions and interactions in the task, in an unobtrusive way. A logfile holds the record of which task is being taken, unique event codes, when the event occurred, identification of the test-taker, and information including identification of the partner in the collaborative dyad, and page of a task on which events occurred (some tasks have multiple pages). It records whether the student is Student A or B of a two student pair, the event and action taken by the student (including chat), start, finish, progress within a task, a description of the action and the text content of the chat. The combination of time and activity provides useful information in its own right as well as data for inferential purposes.

4.3.6. *Coding*

The student pair actions and chats, and their sequences, in the logfile data were analysed to verify the behaviours that could be mapped to specific elements and levels of the elements within the framework. These behaviours were coded into rule-based indicators. Adams, Vista, Scoular, Awwal, Griffin, and Care (2015) describe the approach to coding the logfile data and to scoring in detail. An indicator that was identified as a demonstration of a social skill, for example, was hypothesised to be chat occurring before a particular action. This indicator can determine the frequency and relevance of interaction by a student. The coding for this would identify the specific task, identification of the action-chat sequence, and the specific student (A or B).

From the Laughing Clowns task, an example of an indicator captured from the data and identified as a cognitive skill in the framework is the testing of all positions (L, M, R) by a student provided that the student has access to at least three of the 12 balls. This was hypothesised to indicate a relatively systematic approach to the exploration of the problem. A more strategic approach could be demonstrated by a student when every combination of L, M, R is tested. This would indicate a highly systematic and exhaustive approach to the exploration of the problem space.

4.3.7. *Scoring*

Once the indicators were developed, algorithms were programmed to capture the specific sequence of events in the logfile data. The coded indicators became student data points on variables. Frequency of occurrences of the indicators was interpreted as a proxy measure of difficulty. When these are linked to the difficulty levels in the blueprint, interpretation is possible. Each of the 11 tasks was designed to require, and to indicate, different levels of difficulty across elements. Therefore when several tasks are completed by individuals as a bundle, it provides the capacity for a wider variety of difficulty levels to be catered for. The method for understanding assessment scores in terms of performance criteria with relevant stages of competence was drawn from the work of theorists Glaser and Rasch. Glaser (1963) developed a criterion-referenced approach to the interpretation of assessment scores indicated by behaviours. The student's performance is determined along an increasingly continuum of competence which Rasch's (1980) work enables to be mapped.

5. Results

5.1. *Calibrating*

From the scored data, the ability of each student and the relative difficulty of the indicators was estimated. Table 4 shows the excellent item difficulty and the fit to the

Table 4. Item difficulty and fit to the Rasch model of the Laughing Clowns task.

Strand	Element	Indicator	Difficulty Estimate	Measurement Error	Weighted Mean Square
Part	Interaction	8	0.895	0.039	0.98
Part	Interaction	12	0.424	0.039	1.05
Part	Interaction	4	0.454	0.039	1
Pers	Responsiveness	16	1.464	0.039	1.04
Pers	Responsiveness	17	-2.714	0.045	0.94
SocReg	Responsibility initiative	5	-0.435	0.039	1
SocReg	Responsibility initiative	13	-0.523	0.039	0.98
TR	Problem analysis	3	1.01	0.04	1
TR	Problem analysis	11	1.094	0.04	0.98
TR	Resource management	1	3.106	0.046	1.13
TR	Resource management	9	0.267	0.039	1.06
TR	Collects information	18	-0.646	0.166	0.98
TR	Systematicity	2	-0.686	0.04	0.97
TR	Systematicity	10	-0.657	0.04	1
KB	Reflects and monitors	6	-1.409	0.042	0.98
KB	Reflects and monitors	14	-1.416	0.042	0.97
KB	Solution	7	-0.218	0.039	1.01
KB	Solution	15	-0.011	0.039	0.99

Notes. Confidence interval is constant from 0.91 to 1.09.

TR = Task Regulation, Part = Participation, SocReg = Social Regulation, Pers = Perspective Taking, KB = Knowledge Building

Rasch model for the Laughing Clowns task. The most difficult item (Item 1 at 3.106 logits) is represented by a student using only half the available balls to determine the pattern. Achievement of this item requires a high level of setting goals and resource management combined with high systematicity. Just above the middle of the distribution, and indicating communication through a threshold value for number of chats before all balls are placed immediately prior to action decisions, occurs Item 4 (at 0.404 logits). At the least difficult level, both students come to consensus on the solution (Item 6 at -1.409 logits), sampling the relatively easy activity of negotiation in this instance.

Each of the tasks was calibrated separately, and then calibrated concurrently to identify that they mapped onto the same underlying dimension. Concurrent equating ensured that which set of tasks was undertaken was a matter of indifference because they each contributed information about student ability. Figure 2 depicts the item-person map for the five strands – Participation, Perspective Taking, Social Regulation, Task Regulation and Knowledge Building. On the left hand side of the figure distributions of student ability are shown for each strand, while on the right hand side, item numbers of indicators representing the elements are listed. The quantity of the indicators provides a quick view of the capacity of the tasks to generate sufficient indicators to map student ability across the range for the strands.

5.2. *Estimating student ability*

Each task has different numbers of indicators that can be demonstrated. In addition, not all students go about a task in the same way so there are different indicators and different numbers of indicators demonstrated by each student depending on their approach to the task. Hence the ability estimate is based on which indicators are used by the student and the p = maximum possible indicators, given both the set of tasks attempted and Student A or B identity of the student (noting that in the asymmetric tasks, different opportunities are available to each student). The concurrent task equating provides the assurance that regardless of bundle taken by the students, their results will be equivalent.

The purpose of the assessment and the calibration was to obtain estimates of student ability in each of the five strands. These are then used to locate the student on a level on the developmental progression formed by interpretations of the indicators located across the difficulty range and within strand. Griffin (2007) outlined the theoretical basis for identifying levels, their cut scores, and the subsequent interpretation of levels. The cut scores and clusters of indicators lead to an interpretation of the variable and the definition of the developmental progression. Location of a student on this progression provides the identification of the skills the student is ready to learn. The student location is based on the point at which they have a 50:50 chance of successfully completing an item. An interpretation for the progression for Task Regulation, which was derived in this way, is presented in Table 5. Such empirically-derived progressions make possible the linking of the framework elements with their hypothesised low, middle and high levels to the descriptions on these developmental progressions.

Part	Pers	SR	TR	KB	Item distribution
4					46
					52
					2
					88
3	X				1 5 20
	X				6 55
	X				18 58
	X				89
	XX				
	XXX				
	XXXX				56 71 91
2	XXXX				42 43 74 83 135
	XXXXXX				40 129
X	XXXXXX				35 92 132 144
X	XXXXXXXX	X			126 153
XX	XXXXXXX	X			15 34 47 123 138
XXX	XXXXXXX	X			141 147
1 XXXXX	XXXXX	XXX	X		31 51 68 80 150
XXXXXX	XXXXXXX	XXXX	X		27 63
XXXXXXXX	XXXXXX	XXXXX	XX	X	48 77 85
XXXXXXXXXX	XXXX	XXXXXXX	XXX	X	3 22 45 49 60 73 78
XXXXXXXXX	XXX	XXXXXXX	XXX	X	7 16 39 44 59 66 82
XXXXXXXX	XXX	XXXXXXXX	XXXXX	XX	21 67 70 90
XXXXXXXX	XX	XXXXXXXX	XXXXX	XXX	4 24 50 53 79
0 XXXXXXXX	X	XXXXXXXX	XXXXXX	XXX	19 37 101 102
XXXXXXX	XX	XXXXXXXX	XXXXXX	XXX	57 62 86 106
XXXX	X	XXXXXX	XXXXXXX	XXX	26 54 65 76 99 100 105 134
XXX	X	XXXXX	XXXXXXXXX	XXXX	72 107 110 111 115 116 133
XX		XXXX	XXXXXXX	XXXXX	25 30 33 36 41 95 96 97 108
XX		XXX	XXXXXX	XXXXX	29 64 75 81 87 93 98 117
X		XX	XXXXX	XXXXXX	10 17 23 69 84 103 104 128
-1		X	XXXXX	XXXXXX	13 38 136
		X	XXXX	XXXXXX	28 32 114 121 122 131 154
		X	XXXX	XXXXXX	14 112
		X	XX	XXXXX	118 119 140 146 155
			XX	XXXXX	8 125 142 143 145 149
			X	XXXXX	113 120 127 130
-2			X	XXXX	11 139
			X	XXX	61 151
			X	XX	124 148
				XX	152
				XX	
				XX	9
				X	
-3				X	
					94
					12
					109
-4					

Each 'X' represents 45.8 cases

Note. Part = Participation; Pers = Perspective Taking; SR = Social Regulation; TR = Task Regulation; KB = Knowledge Building; adapted from Griffin and Care (2015)

Figure 2. Concurrent calibration variable map for five strands.

Table 5. Levels in the developmental progression for the Task Regulation strand.

Level		Description
High	F	The student's approach to the task is systematic and they work very efficiently, successfully completing complex tasks in an optimal amount of time and attempts. They work with their partner to identify the relevant resources and disregard those that posed no benefit in previous attempts.
	E	The student's engagement in the task appears to be well thought out and planned and each action appears purposeful. The student plans goals based on knowledge and experience from previous goal outcomes and subtask completion. They note information that may be useful in future tasks or for an alternative solution path.
	D	The student adopts strategic sequential trials and increasing systematic exploration. They narrow their goal setting and focus on successfully completing a subtask before moving on. The student simplifies the problem, analyses it in stages and plans strategies with their partner.
	C	The student becomes aware of the need for more information pertaining to the task and begins to gather as much information as possible. The student realises that they may not have all the required resources and allocate their own resources to their partner.
	B	The student limits their analysis of the problem by only using the resources and information they have. They make good use of their own resources. The student will remain limited in their goal setting with broad goals such as completing the task.
Low	A	The student attempts to solve the problem through a random guessing approach and tends to repeat previous mistakes or trial the same approach multiple times with no clear indication of advancing through the task. However, if the student has difficulty in understanding the task they make very little attempt to explore the problem at all.

5.3. *Interpreting indicator locations*

Interpreting the locations of the indicators in the context of mapping against the blueprint informed the development of the reporting module. Consistent with the notion of skills development and use of the zone of proximal development (Vygotsky, 1978) as facilitating learning and teaching, scores against collaborative problem solving are not reported. Instead, the identification of the student at a point along the progression annotated by the skills demonstrated is presented. The summary statements for each level are shown in the reporting module. Based on the student ability estimates being derived as outlined here, reports can be generated both for an individual student and for a whole class.

6. Conclusion

Collaborative learning refers to students working together to achieve a common goal in a shared learning environment (Underwood & Underwood, 1999; O'Neil et al., 2003). In collaborative problem solving however, that common goal becomes a problem that the group needs to work together to solve.

One main difference between collaborative learning and collaborative problem solving is in the nature of the activity. Much of the support for the importance of working collaboratively comes from social constructivists such as Vygotsky (1978) who suggested that social interaction facilitates learning. Therefore, generally, the nature of collaborative learning ensures students are provided with well-defined activities, explicitly informed of the goal and instructed to work together under the assumption that it will improve their learning and understanding. While some collaborative learning models introduce cognitive elements into the process such as decision making, the focus of collaborative learning is on the ability to learn from the interactive situation (O'Neil et al., 2003).

The nature of collaborative problem solving introduces additional cognitive processes to the equation of working collaboratively. It focuses on how cognitive processes such as goal setting, connecting information and patterns and testing hypotheses can be managed in a collaborative environment. Collaborative problem solving may be the strategic choice for problems that are beyond the resources of any one person to solve. Such problems require that individuals work together as they cannot complete the activity alone. There is mutual benefit from working together when the problem is a common one; such a situation provides the motivation for individuals to participate and work together. Collaborative problem solving activities at their novice levels need to lead to activities at more sophisticated levels which will align with real-world problems. These latter problems are frequently ill-defined – which is why they ARE problems. In these situations, only partial information or resources are typically available, and the skill of the collaborators lies in the cognitive processes they bring to this space and how they coordinate their activities. Building these skills at novice levels can be enabled by artificially with-holding resources from collaborating partners such that only partial information and resources are available to each. The activity itself can be at a level of complexity where identifying the goal may not be straightforward and where the dynamic process of problem solving might prompt alterations in strategy and approaches. The skill set of each participant is also recognised as different, with each individual bringing their own capacities and perspectives to the problem. Through the scaffolding that can be provided by the artificial task environment, the collaborating partners are provided with the opportunity to learn and to build knowledge through the task.

These characteristics of collaborative problem solving stimulate the need for new approaches to assessment, as exemplified in this paper. The development of assessment tasks was a step toward constructing an online environment or problem space that would elicit the skills of interest in such a way that students would learn from and about the process and would become self-conscious about that process, and that teachers would be able to understand the level at which students were functioning, and would be provided with insights about how they might construct classroom-based tasks that would similarly elicit and stimulate the skills. In the future there is a need to move away from an idiosyncratic approach to construction of tasks, and to develop a template approach that will provide efficiencies for large scale assessment, as well as making the resources more easily available for teachers to use in the classroom.

Another approach to the definition, description, and assessment of collaborative problem solving has been taken by the OECD (Greiff, 2013) for the 2015 PISA study. The OECD definition includes the individual problem solving processes that framed the PISA assessments over the past cycles in 2008 and 2012. These include Exploring and Understanding, Representing and Formulating, Planning and Executing, and Monitoring and Reflecting. In addition, social processes of establishing and maintaining shared understanding (knowledge flow/resources), taking appropriate action to solve the problem (task behaviour), and establishing and maintaining team organisation (organisation/management), are considered. These three latter social function sets incorporate many of the same elements identified by Hesse et al. (2015), but set these in a matrix against the cognitive functions, and at a more general level than the Hesse et al. approach. The major difference between the two conceptual frameworks lies in the PISA functional process approach versus the Hesse et al. cognitive and social psychology construct approach. In addition, the assessment approach described here has as its focus the capture of the process rather than the solution, taking into consideration the reality that novice learners might solve the same problem as more advanced learners, but using more simple processes. The advanced problem solver needs to acquire more advanced processes in order to bring these to bear on complex problems when the simple approaches are not adequate.

Another difference between the approach described here and the OECD approach lies in the structuring of their problem space where the OECD team has designed an environment in which students respond to pre-slugged responses and stimuli represented by an avatar. There are obvious advantages for automation and ease of scoring in taking this approach. A disadvantage may lie in the scaffolding of students through problem solving processes such that they are not able to learn and build knowledge, as can occur in a more ambiguous environment which reflects real life problems. In this article we have outlined an approach to assessment that can be implemented in the less scaffolded environment that is a natural characteristic of a problem solving space which does not value correct solutions over cognitive and social processes.

The identification of a construct, exploring its theoretical underpinnings, and describing what the latent variable might look like in practice, is an unexceptional process. In our treatment of collaborative problem solving, development of understanding of the construct occurred coincident with exploration of its measurement. Due to the complex nature of the construct, its relative ambiguity, its identification as a 21st century skill, and the view of the research team that virtual communications and problem solving are essential for student skill development, the task development process is as complex as the construct. The idiosyncrasies of the approach are due to the innovative nature of the assessments. The validation of the approach has occurred throughout the process, through face validity exercises with teachers and students, and then through statistical analysis of coded data and its modelling. There is no doubt that the approach works, but it is unwieldy. In order to streamline the process of developing measures of collaborative problem solving and other 21st century skills, a template or model approach needs to be designed. This will make assessments both in their creation and their coding more efficient, and will also provide teachers with more clear illustrations of the types of activities they can present to students to enhance their skill development.

The approach to assessment outlined here is amenable to implementation in schools. The approach need not rely on digital networks, but can be integrated into teachers' current practices through their understanding of the complexities of the skillsets. Teaching within traditional curricular areas can incorporate these skills through teachers' manipulation of required resources and pedagogical strategies. Of course, the mature implementation of the skills is assumed to occur increasingly in the digital world, but in its early skill development can be seen in face to face activities in the classroom. In keeping with a view of 21st century education outlined by Griffin et al. (2012), 21st century education needs to be knowledge-centred, so that students can move beyond current understandings to negotiate lives in a dynamically changing world; it needs to be learner-centred with students actively engaged; community-centred so that knowledge building is collaborative; and assessment-centred so that progress can be monitored.

References

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

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). Dordrecht: Springer.

Csapó, B., Ainley, J., Bennett, R., Latour, T., & Law, N. (2012). Technological issues for computer-based assessment. In P. Griffin, B. McGaw & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 143–230). Dordrecht: Springer.

Darling-Hammond, L. (2012). Policy frameworks for new assessments. In P. Griffin, B. McGaw & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 301–340). Dordrecht: Springer.

Dehler, J., Bodemer, D., Buder, J., & Hesse, F. W. (2011). Guiding knowledge communication in CSCL via group knowledge awareness. *Computers in Human Behavior*, *27*(3), 1068–1078.

Delors, J., Al Mufti, I., Amagai, I. A., Carneiro, R., Chung, F., Geremek, B., . . . Nanzhao, Z. (1996). *Learning: The treasure within*. Paris: UNESCO.

Fischer, A., Greiff, S., & Funke, J. (2012). The process of solving complex problems. *The Journal of Problem Solving*, *4*(1), 3.

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes: Some questions. *American Psychologist*, *18*, 519–521.

Gordon, J., Halasz, G., Krawczyk, M., Leney, T., Michel, A., Pepper, D., . . . Wisniewski, J. (2009). *Key competences in Europe: Opening doors for lifelong learners across the school curriculum and teacher education*. CASE Network Report No. 87. Warsaw: Center for Social and Economic Research.

Greiff, S. (2013). Interactive and collaborative problem solving. *Review of Psychology*, *19*(2), 111–121.

Griffin P. (2007). The comfort of competence and the uncertainty of assessment. *Studies in Educational Evaluation*, *33*(1), 87–99.

Griffin, P. (2014). Performance assessment of higher order thinking. *Journal of Applied Measurement*, *15*(1), 53–68.

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

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

Hesse, F., Care, E., Buder, J., Sassenberg, K., & Griffin, P. (2015). A framework for teachable collaborative problem solving skills. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills: Methods and approach*. Dordrecht: Springer.

O'Neil, H. F. (1999). Perspectives on computer-based performance assessment of problem solving. *Computers in Human Behavior*, *15*, 255–268.

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. (2004). *The definition and selection of key competences: Executive summary*. OECD.

OECD. (2013). *PISA 2015: Draft collaborative problem solving framework*. http://www.oecd.org/pisa/pisaproducts/

Peterson, R. S., & Behfar, K. J. (2005). Leadership as group regulation. In D. M. Messick & R. M. Kramer (Eds.), *The psychology of leadership: New perspectives and research* (pp. 143–162). Mahwah: Erlbaum.

Pollastri, A., Epstein, L., Heath, G., & Ablon, S. (2013). The collaborative problem solving approach: Outcomes across settings. *Harvard Review of Psychiatry*, *21*(4), 188–199.

Polya, G. (1973). *How to solve it: A new aspect of mathematical method*. Princeton, NJ: Princeton University Press.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago: The University of Chicago Press.

Rychen, D. S., & Salganik, L. H. (Eds.) (2003). *Key competencies for a successful life and a well-functioning society*. Cambridge, MA ; Toronto: Hogrefe & Huber.

Scardamalia, M., Bransford, J., Kozma. R., & Quellmalz, E. (2012). New assessments and environments for knowledge building. In P. Griffin, B. McGaw & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 67–142). Dordrecht, Germany: Springer.

Underwood, J., & Underwood, G. (1999). Task effects on co-operative and collaborative learning with computers. In K. Littleton & P. Light (Eds.), *Learning with computers* (pp. 10–23). New York: Routledge.

Vygotsky, L. (1978). Interaction between learning and development. In *Mind and society* (pp. 79–91). Cambridge, MA: Harvard University Press.

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Erlbaum.

Wilson, M., Bejar, I., Scalise, K., Templin, J., Wiliam, D., & Torres-Irribarra, D. (2012). Perspectives on methodological issues. In P. Griffin, B. McGaw & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 67–142). Dordrecht, Germany: Springer.