Educational Assessment in an Information Age

Patrick Griffin
Esther Care *Editors*

Assessment and Teaching of 21st Century Skills

Methods and Approach

Springer

Educational Assessment in an Information Age

Series Editors

Patrick Griffin
Esther Care

This series introduces, defines, describes and explores methods of developing new assessment tools in educational and work environments which are increasingly characterised by use of digital technologies. Digital technologies simultaneously demand, reflect, and build student skills in many areas of learning, old and new. They can and will continue to adapt and facilitate the assessment of both traditional academic disciplines as well as those known as 21st century skills. These skills include creativity, critical thinking and problem solving, collaborative skills, information technology skills, and new forms of literacy, and social, cultural, and metacognitive awareness. The capacity of digital technologies to capture student learning as a process as well as student achievement is vast. New methods need to be developed to harness this capacity in a manner that can produce useful and accurate information to teachers for classroom interventions, and to education systems for policy development. The series includes innovative approaches to assessment in terms of their psychometrics and technology platforms; outcomes of implementation of assessments of generic skills at large scale in the classroom; and use of large scale assessment data to inform policy in education. The series explores the assessment of new and emerging skills required of graduates and how new forms of assessment inform teaching; it projects into the future the kinds of assessment possibility associated with technology and explores the assessment links between education and the workplace.

More information about this series at http://www.springer.com/series/13475

Patrick Griffin • Esther Care
Editors

Assessment and Teaching of 21st Century Skills

Methods and Approach

Springer

Editors
Patrick Griffin
Esther Care
Assessment Research Centre
Melbourne Graduate School of Education
University of Melbourne
Parkville, VIC, Australia

ISBN 978-94-017-9394-0 ISBN 978-94-017-9395-7 (eBook)
DOI 10.1007/978-94-017-9395-7
Springer Dordrecht Heidelberg New York London

Library of Congress Control Number: 2014952887

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

This book is the second volume in a series dealing with the identification, development, measurement and teaching of 21st century skills. The first volume was an edited collection of white papers that established a conceptual basis for the Assessment and Teaching of 21st Century Skills project. This volume explores the methodologies that underpin the work of a large team of people from six countries in developing and implementing the ideas of the project. The third volume will be a collection of research papers reporting on empirical studies into the reliability, validity and applications of the 21st century skills assessment.

The purpose of this volume is to make transparent the procedures and methodologies, the thinking and strategies, the data and analyses of ATC21S so that readers can understand the application of the theory that went into the project. We also hope that other researchers and measurement specialists will be encouraged to take up any opportunities or challenges raised by our methodology and find ways to improve the process, the products and deliverables. But most of all we hope that these researchers and specialists find ways to help teachers improve student learning in these important 21st century skills.

The idea of formative assessment has been central to the ATC21S project since its inception. The aim has been to inform teaching and ensure that teachers have both the information and the resources to enable them to build and develop 21st century skills among the students they teach. Formative assessment is about informing the teacher. We have used the phrase *informative assessment* at times in order to ensure that the data that are generated in our assessments can be used by teachers to identify student readiness to learn. More importantly, we hope that teachers using the data will change their own practices in ways that accelerate the growth and development of their students. This makes the project different from most other large scale projects which focus on 21st century skills.

The ATC21S project (www.atc21s.org) took as its first aim the definition of 21st century skills, characterising them as skills that are essential for navigating education and the workplace in the current century. A primary characteristic of these skills is that they involve a need for, manipulation of and use of information. The perspective taken by the project was that the identified skills did not need to be new.

Rather, it was argued, 21st century skills are those that must be brought to bear in today's worlds of education and employment in order for individuals to function effectively as students, workers and citizens. Given the recent rapid global changes in technology, certain of these 21st century skills will be new, while others may be traditional skills that need some adaptation for their implementation. Binkley et al. (2012) described recent changes in the knowledge economy under four headings: Ways of Thinking, Ways of Working, Tools for Working, and Living in the World. Collaborative problem solving operates at the intersection of critical thinking, problem solving, communication and collaboration in the framework of the ATC21S project (Griffin et al. 2012). The second area of development, Learning through Digital Networks (LDN-ICT) is introduced in Chap. 3 and will be extensively operationalised in a later volume.

Notwithstanding the process-orientation of these themes, the substance upon which processes are enacted is acknowledged and emphasised in ATC21S through recognition of the need for discipline-based knowledge and understanding. Processes are best learned in application: although the need for skills such as critical thinking and problem solving is often recognised, teaching those skills is often divorced from content. Educational programs need to integrate the development of skills within the learning of content. We use the word "development" deliberately, because we are describing a process that students progress through in a developmental manner via stages of increasing competence. A challenge for the identification, description, assessment and teaching of 21st century skills exists in bringing about acceptance of the concept of new literacies which might inform learning and teaching. In the same way that language literacy and numbers literacy are accepted as requirements for learning and teaching in language and numeracy, ICT skills and problem solving approaches can be viewed as forms of literacy. In the ATC21S project, the development of assessment tasks has rested on the idea that while the skills may be generic, their utility is demonstrated through students developing and using them in their curricular studies.

In the project, 21st century skills were defined as activities where groups execute a number of steps in order to transform a current state into a desired goal state, or to move through cognitive states from analysis of information to hypothesis testing. In defining collaborative problem solving, Hesse, Care, Buder, Sassenberg and Griffin (2015, this volume) describe both social and cognitive components. To solve a problem, it may be that a variety of content knowledge, different strategies, different resources, or different skills are required, and that not all of these are possessed by one individual. When a task has this level of complexity, identification of the problem or of the goal itself may present challenges, both in the way a group of individuals approaches the task and in the choice of processes used by the individuals to solve it. In 2015, the OECD has elected to use interactions between humans and computers (human to computer-agent, or H2A) as a means of educational measurement, whereas Griffin and Care (2015) in the ATC21S project have elected to use interactions between humans (human to human, or H2H) within a technology medium. The effectiveness of these approaches and their validity remains to be resolved. The discussion in this volume does not enter into the debate about the medium of presentation of the assessment tasks. This volume is about the methodology of the ATC21S project. Further discussion on the medium of task presentation and the roles of agents or scripts in a collaborative environment will be included in Volume 3.

This book and its research project are set in the context of recognition that education is changing. Education needs to prepare students to deal with rapid changes in employment and learning styles. Teachers need to prepare students for jobs that have not yet been created. In the future there will be technologies that have not yet been invented; there will be ways of living and thinking and learning that have not yet emerged. Students will need to leave school and universities with skills, attitudes and values more commensurate with a digital information age. Education is now about the preparation of students for new ways of thinking: ways that involve creativity, critical analysis, problem solving and decision making. Students need to be prepared for new ways of working that will call upon their communication and collaboration skills. They will need to have a familiarity with new tools that include the capacity to recognise and exploit the potential of new technologies. In addition they will need to learn to live in this multifaceted new world as active and responsible global citizens.

The employment that these students are likely to enter will increasingly require critical and expert thinking skills and complex forms of communication. It is for most countries a formidable economic problem to prepare graduates for this new kind of workforce. Those wishing to be highly rewarded in the workforce of the future will need to be expert at interacting with people to acquire information, to understand what that information means, and to persuade others of its implications for action. This is not to say that the foundation skills of numeracy and literacy are becoming irrelevant. New forms of numeracy and new forms of literacy will emerge, but the capacity of an individual to work with numerical data, to access and interpret and use information from a number of sources, will always remain important. Problem solving skills in a technology-rich environment will become increasingly important. So, too, will the capacity to communicate, collaborate and create. As the globalised world becomes more complex and integrated across national boundaries, individuals will need to be able to cross those boundaries to collaborate on shared information and emerging knowledge. The more complex the world becomes, the more individuals will need these skills. The more content can be searched and accessed, the more important the filters and explainers will become – they will need to be able to build problem solutions by identifying components and linking them together in ways that make sense to them and to other people. It is also the case that in the 21st century the idea of fixed employment for 30 or 40 years has disappeared. In the 21st century, students leaving school or university can expect to have 10–15 different jobs in their work life. In order to successfully enter this new workforce, these people will need to have a new breadth and depth of understanding and a capacity to learn and re-learn. They will not be masters of one particular field, but will have the capacity to learn and adapt across many fields in their work life.

For all these reasons, the importance of the ATC21S project cannot be overstated. This book will, we hope, help academics, policymakers and teachers, as well as parents and industrialists, in adapting their workforce, their graduates, their friends and their associates to work and live more successfully in the new digital world.

The book contains 15 chapters in five parts. Part I deals with the overview of the project. It contains Chap. 1 which provides an overview of the methodology and the nature of the project. The chapter details the method of the project, focusing on the development and calibration of the collaborative problem solving tasks. The method is important because in this project we have undertaken to be transparent in order to enable others to take what we have done and improve on it.

Part II deals with the conceptual nature of the project and the measures it derived. It contains Chaps. 2 and 3. In this part we define the conceptual basis of the 21st century skills that were chosen for development in this project. Chapter 2 presents the conceptual framework for collaborative problem solving and Chap. 3 describes the learning through digital networks construct both conceptually and through some empirical data.

Part III deals with the technical aspects of the development of the assessment tasks for collaborative problem solving. It contains a series of chapters that describe the delivery and interpretive mechanisms employed in the project. This includes the platform that needed to be developed to enable collaborative work to be undertaken on the Internet. This platform is explained in Chap. 5. One of the innovative aspects of the project was the automatic coding and interpreting of log stream data. This is explained and illustrated in Chap. 6. The data from this coding and analysis enabled more sophisticated work with item response modelling to be applied to the data. How this was applied to each of the tasks is explained and documented in Chap. 7.

Part IV deals with fieldwork aspects of the task development process. The chapters in this part discuss the work implemented in the six participating countries. Chapter 8 focuses on the work undertaken in Australia; Chap. 9 on the processes in Singapore; Chap. 10 on Finland; Chap. 11 on USA; Chap. 12 on Costa Rica; and Chap. 13 on the Netherlands. Each takes a slightly different approach and together they illustrate how the same core processes vary across countries according to their different needs and imperatives.

Part V addresses the implications of the project for educational issues that arise in the classroom and at the system level. In Chap. 14 suggestions are made about how the work of the project can be used in the classroom. In Chap. 15 the authors describe procedures that may help jurisdictions and education systems to take this project to scale.

The book is written for a number of audiences. Firstly, measurement people: we hope that our presentation of the measurement procedures introduced and followed in this project will enable others to improve, refine and critique the procedures and the analysis that we have used. Our procedures follow closely the chapter by Wilson et al. in Volume 1. The scoring procedures described in Chap. 6 comprise a very complicated process. It is our wish that others will find a way to simplify this process. The authors of this volume who are members of the University of Melbourne team will also be working in the near future to provide templates and simplified coding and scoring procedures. We hope it will be possible to include these in Volume 3.

A second audience for whom this volume is written comprises policymakers, education ministers, permanent secretaries, classroom teachers and postgraduate students. Finally, the book is about documenting procedures and data. It is about transparency. The debates that have emerged since this project began – debates about the way in which 21st century skills, such as collaborative problem solving, can be presented to students in an assessment format – have not been entered into in this volume. The authors of this volume are unanimous in their view that collaborative work involves people-to-people interaction and that people interacting with a computer does not amount to collaboration.

There are several innovative characteristics of this project and this volume sets out to detail each of them.

1. Most research on teamwork has focused on the outcomes of entire groups. Individual assessments are normally conducted externally, in isolation from the collaborative task. Through this approach it has not been possible to estimate individual group members' skills as demonstrated during the task: there has been a lack of opportunity or process to analyse interactions during the collaborations. The current volume addresses this problem by showing how the collaborative efforts of individuals can be measured.
2. The role of education technology in 21st century assessment enables computers to provide detailed time-stamped data capturing the activities of collaborators. The resulting activity logs provide log stream and chat stream data for modelling and evaluating student activity.
3. The study brings together a focus on assessment in ways that are unusual and seldom implemented. It follows that the ATC21S project has pioneered the H2H approach.
4. The study focuses on the development of problem solving within an inductive-deductive paradigm. The volume explores how this hierarchy is supported within a collaborative context and what implications it has for teaching hypothetico-deductive reasoning skills in two diverse curriculum areas.
5. A great deal of the work reported in this volume has been both pragmatic and conceptual. Discussions regarding the effects of mixed ability collaborations, role allocation and the call for empirical investigations of these issues have given this project an important base. We hope that its challenges will be taken up and that the cutting edge of 21st century assessment will be pushed even further into the future.

Parkville, Australia Patrick Griffin

References

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.

Contents

Part I
Project Method and Overview

In this part, Griffin and Care (2015, Chap. 1) explain the overall plan of the project. The part shows how the project follows methods normally used in test construction for large-scale survey testing. In building complex assessment tasks, it was essential to develop conceptual frameworks for each of the constructs being measured. Expert groups were formed in order to define the frameworks and the constructs and each expert group contained at least one measurement specialist to ensure that the constructs had essential measurement properties such as evidence of order, direction and magnitude and that the constructs were able to be taught and leaned. From these deliberations, the project team developed blueprints for assessment construction, describing the skills needed by students undertaking the collaborative problem solving and learning through digital network tasks. The project team commissioned three organisations to draft assessment tasks which would map onto the hypothesised constructs. The commissioned developers were World Class Arena in Britain, University of California (Berkeley) and the University of Melbourne. A series of iterative procedures, checks and monitoring steps were then implemented to ensure that the tasks met criteria established by the project directorate. Once the draft tasks were prepared, they were reviewed by a series of panels as described in Chap. 1. The first panel consisted of teachers and national project managers. They checked that the tasks would be engaging for students, that the tasks could be used for both assessment and instructional purposes; discriminate between high and low performers; and yield sufficient data points for scoring and reporting purposes. The draft tasks were then subject to cognitive laboratory procedures in which the national project managers and other field workers observed the students undertaking the tasks and encouraged the students to 'think aloud' in order for the research team to understand the thinking processes that students used while solving the tasks. These procedures helped the project team develop scoring criteria for the tasks. The assessment tasks were then loaded onto a collaboration enabling platform. Reporting modules were produced, based on previous work of the Assessment Research Centre. A series of professional development modules are also described.

Griffin, P., & Care, E. (2015). The ATC21S method. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills: Methods and approach* (pp. 3–33). Dordrecht: Springer.

Chapter 1
The ATC21S Method

Patrick Griffin and Esther Care

Abstract The ATC21S™ project followed a research and development plan that consisted of five phases: conceptualisation, hypothesis formulation, development, calibration and dissemination. (The acronym ATC21S™ has been globally trademarked. For purposes of simplicity the acronym is presented throughout the chapter as ATC21S.) Within the conceptualisation phase, the project focused on the definition of twenty-first century skills. This chapter outlines the selection and conceptualisation of the skills to be assessed. It describes how this led to the development of hypothesised learning progressions which portrayed how the skills might vary across more and less adept individuals. Assessment tasks were commissioned to be developed from a mixture of commercial agencies and universities. The tasks were then subjected to concept checking, cognitive laboratories, pilot studies and calibration trials. The final stage of the process is dissemination, which includes the development of scoring, reporting and teaching in the Assessment and Teaching of 21st Century Skills system.

Background

In 2008, three large technology corporations (Cisco, Intel and Microsoft) became concerned about the skills of students graduating from school and university. The three companies were alarmed that graduates were entering the workforce with skills that did not prepare them for employment in a digital age. They identified a need to focus on twenty-first century skills because of shifting workplace requirements, as outlined by Griffin et al. (2012a). They reasoned in their 'call to action' that there was an emerging need to change the bases for hiring and firing employees to reflect the possession of twenty-first century skills, and structural unemployment situations in many developed countries seemed to support the call for change.

P. Griffin (✉) • E. Care
Assessment Research Centre, Melbourne Graduate School of Education,
University of Melbourne, Parkville, VIC, Australia
e-mail: p.griffin@unimelb.edu.au; e.care@unimelb.edu.au

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_1

The call was for the identification of relevant twenty-first century skills and the development of means to assess and teach them. The skills needed to be relevant in the twenty-first century but not exclusively so, since many twentieth century skills would remain important.

Dr Robert Kozma was commissioned to formulate a call to action (see Kozma 2009). Subsequently, research was undertaken to identify a leader for a project that would target assessment and teaching with a focus on twenty-first century skills. Professor Barry McGaw was identified as an appropriate leader and he directed the project until the end of 2009, when the lead passed to Professor Patrick Griffin. The project was funded jointly by the three corporations and by six participating governments (Australia, Singapore, USA, Costa Rica, Netherlands and Finland). The project delivered eleven human to human collaborative (H2H) problem solving tasks and two collaborative ICT literacy tasks by the close of the project in July 2012. At that stage all tasks were prepared in Flash. Subsequently materials were made available through the project website ATC21S.org. In addition to the tasks and their interpretation the project prepared a series of professional development modules and student and school reporting modules. These additional parts of the project, together with the reprogramming were all undertaken after the cessation of the project and were funded by the University of Melbourne and its Assessment Research Centre. The corporate funding was essential for the development years (2009–2012), but after July 2012, the Assessment Research Centre(ARC) financed all further work. Following a failed outsourcing of reprogramming and migration to the cloud the ARC also redeveloped all tasks and procedures in HTML5 as a portable system available to members of a global research consortium. This process took more than two years.

The project had three broad goals:

1. To establish baselines and methodologies for the assessment and teaching of the 21st century skills needed in a digital age;
2. To influence development of curriculum and resources so that education could change and move towards more relevance in a digital age;
3. To use the substantial facilities and resources of the corporate sponsors to generate the interest and engagement of countries and academics by fostering new methods of assessment, teaching and learning suitable to the use of digital resources in education curricula.

It was argued that traditional forms of assessment may not be suited to the measurement of many twenty-first century skills, especially those that might be considered non-cognitive, so a precise goal was established: to develop new assessment approaches matched to twenty-first century skills, and to advise systems, schools and teachers on the use of assessment data to help students develop these higher-order skills.

In keeping with this goal, a large-scale discussion seminar and symposium was conducted at the end of the 2009 annual meeting of the American Educational

Research Association in San Diego. More than 250 academics and industry representatives attended the workshop and teams were established to explore five questions:

1. What are the twenty-first century skills and what kind of capacities would it be possible for schools to generate among graduates?
2. What methodologies of assessment and measurement might be appropriate to the assessment of new kinds of skills suitable for the twenty-first century?
3. What kinds of technologies might be used by schools for this purpose? What kinds could be seen emerging in education curricula and how were they already being deployed?
4. What kinds of new teaching approaches and information sharing, knowledge generation and networking within classrooms might facilitate the learning of twenty-first century skills?
5. What factors might be taken into account in bringing changes in education curriculum to scale, and how might this influence policy in education systems?

Five teams of people were established to write 'white papers' associated with these questions. Some discussion ensued regarding the question of whether they identified skills or competencies. In this volume we regard skills as the things people can do and competence as a measure of how well they do them. A competent person adjusts the performance of a skill to the demands of the context in which the skill is required. No one performs at their maximum all the time but can be expected to adjust performances to expectations. We do not know what the typical performance is, but we assess people to determine their maximum competence in order to identify what their potential performance might be. The teams and their leaders were:

- The defining of twenty-first century skills led by Senta Raizen
- Methodological issues led by Mark Wilson, University of California, Berkeley
- Technological issues led by Beno Csapo, University of Szeged, Hungary
- Classrooms and formative evaluation and knowledge generation led by John Bransford, University of Washington, and Marlene Scardamalia, University of Toronto
- Policy frameworks and new assessments led by Linda Darling-Hammond, Stanford University

The white papers were published in Volume 1 of this series (Griffin et al. 2012b) together with an overview document which explored the changing role of the workplace and its relationship to education.

At the same time, while these issues were being explored and documented from an academic point of view, the companies and founding countries established an executive board to assist in the development of new techniques for assessment and teaching of 21st century skills. The members were:

- Barry McGaw, University of Melbourne (member until January 2010)
- Patrick Griffin, University of Melbourne, Executive Director ATC21S (from January 2010)

- Michael Stevenson, Cisco Vice President – Global Education, ATC21S Board Chair (2009–2010)
- Anthony Salcito, Microsoft Vice President – Education, ATC21S Board Chair (2010–2011)
- Shelly Esque, Intel Vice President – Legal and Corporate Affairs, ATC21S Board Chair (2011–2012)
- Esther Care, University of Melbourne, ATC21S International Research Coordinator (from March 2010)
- Ministers of Education or their representatives from the founding countries.

An advisory board was also established. The membership of the advisory board included:

- Patrick Griffin, University of Melbourne, Executive Director ATC21S (Chair)
- Andreas Schleicher, Organization for Economic Co-operation and Development (OECD)
- Seamus Hegarty, International Association for the Evaluation of Educational Achievement (IEA)
- Irina Bokova (Director General UNESCO)
- Ray Adams, Technical Director PISA 2003–2012
- Marc Durando, European Schoolnet
- Esther Care, University of Melbourne, ATC21S International Research Coordinator
- Stuart Elliott, National Academy of Sciences
- David Forster, International Testing Commission
- Robin Horn, World Bank
- Eugenio Eduardo Severin, Inter-American Development Bank

For each participating country, a national project manager was appointed. These national managers also became members of the advisory board. A project task force was established. This consisted of the Executive Director, International Research Coordinator and an executive member of each of the three corporations.

The executive board made decisions about funding, timelines, approaches and the approval of research strategies. The advisory board provided support and assistance in gaining access to large and influential organizations associated with educational assessment and teaching. The task force helped the Executive Director on day-to-day matters of project development and communications.

The project adopted three broad strategies. The first was to influence ministers of education and education systems throughout the world, persuading them that curriculum needed to change in order to make people work-ready in a digital age. The second was to influence other corporations and large-scale employers of school-leavers and graduates of universities, persuading them to shift employment criteria to a greater focus on twenty-first century skills, rather than the outcomes of nineteenth or twentieth century education. The third was to influence major educational monitoring and evaluation organizations, such as OECD through its PISA project, and the IEA through its influence in cross-national studies.

The Framework

The working party focusing on the definition of twenty-first century skills argued that these skills were underpinned by a set of knowledge, skills, attitudes, values and ethics. This conception of twenty-first century skills became known as the KSAVE model. The KSAVE model was treated as an overarching framework within which the skills and developing requirements of the twenty first century were considered by the five white paper writing teams. Raizen's team focused on four broad categories of skills (see Binkley et al. 2012):

- Ways of thinking: Creativity, *critical thinking*, *problem solving, decision-making*, learning and innovation
- Ways of working: Communication and *collaboration*
- Tools for working: *Information and communications technology (ICT)* and *information literacy*
- Living in the world: Citizenship, life and career, and *personal and social responsibility*

The set of skills considered in the white papers did not directly address attitudes, values or ethics. Instead, the initial focus addressed skills and resolved to explore other aspects of the KSAVE framework later.

Selecting the Skills

In January 2010 project leaders and directors met in London and identified three broad skills that could be assessed as twenty-first century skills. They also explored the possibility that the selected twenty-first century skills could be taught and learned. They reasoned that the combination of critical thinking, problem solving, decision-making and collaboration could be conflated into a single complex set of tasks or skills under the title "Collaborative Problem Solving". They further reasoned that information literacy, information and communications technology literacy and personal and social responsibility could be conflated into the ways in which students learn through social networks and social media. The title later given to this set of skills was "Learning through Digital Networks". In taking this approach, the project was able to address more than half of the twenty-first century Skills defined by Raizen's team.

Expert Panels

Three types of expert panels were formed with the cooperation of the national project managers.

Substantive Expert Panels

The initial panels consisted of specialists or researchers in content or a related area with a measurement advisory member. The task was to ensure that the hypothesised constructs were theoretically sound, that the evidence could be mapped onto a developmental progression, and that a framework could be established within which the assessment tasks and teaching strategies could be conceptually based. In order to define the skills more fully, conceptual frameworks were developed. Wilson and Scalise (2015; Chap. 3) outline the membership of the team and the conceptual framework for learning through digital networks. A team led by Friedrich Hesse explored and defined the components of collaborative problem solving. Each of the expert panel leaders was assigned specific roles.

Roles and responsibilities of the Expert Panel Leaders were:

1. Attend combined three day Panel meeting
2. Convene and coordinate the work of the Panel
3. Be familiar with all the White Papers in order to contextualise the task
4. Provide a theoretical rationale for the hierarchical content of the developmental learning progression
5. Collate materials related to the development of the hypothesised learning progression
6. Draft a minimum of three levels in a hypothesised developmental learning progression
7. Direct and lead the search of task banks and the selection of tasks matching the developmental progression
8. Identify gaps in available assessment materials
9. Liaise with International Research Coordinator
10. Assemble materials for delivery to the post AERA conference meeting in Denver in 2010
11. Provide advice to the Project Director on possible task development agencies
12. Produce a report regarding issues and problems to be resolved concerning the assessment of the twenty-first century skill nominated by mid 2010

The role of the Panel Members was:

1. Attend a three day Expert Panel meeting
2. Contribute to the following activities :
   (a) Draft hypotheses regarding the developmental learning progression
   (b) Identify tasks from relevant task banks
   (c) Nominate the domain space for the development of additional tasks
   (d) Contribute to the preparation of a panel report

Learning Through Digital Networks

The conceptual frameworks underpinning learning through digital networks are described, highlighting the development through four threads (Wilson and Scalise 2015):

(i) Functioning as a Consumer in Networks involves obtaining, managing and using information and knowledge from experts and shared digital resources in order to benefit private and professional lives
(ii) Functioning as a Producer in Networks involves creating, developing, organizing and reorganizing information/knowledge in order to contribute to shared digital resources
(iii) Participating in the Development of Social Capital through Networks involves using, developing, moderating, leading and brokering connectivities within and between social groups in order to marshal collaborative action, build communities, maintain an awareness of opportunities and integrate diverse perspectives at community, societal and global levels
(iv) Participating in Intellectual Capital (collective intelligence) in Networks involves understanding how tools, media and social networks operate and using appropriate techniques for operating those resources to build collective intelligence and integrate new insights into personal understandings

The threads were conceptualized as interconnected developments and expanded with hypothesised progress maps and a hierarchy of skills and competencies. Wilson and Scalise (2015) provide a description of the hierarchy of skills and competencies for each thread, for which there are specific competencies identified for a user functioning at low, mid or high levels. This structure assumes that a person who exhibits competencies at a higher level (someone who would be considered highly literate in ICT), will also exhibit competencies at lower levels in the same thread (a novice or beginner). Note that the threads are not mapped at the same fixed level. Hence it is possible that a certain skill level in one thread may be an essential but not contributing factor to the development of skill within another thread.

Collaborative Problem Solving

Friedrich Hesse and his team (Hesse et al. 2015; Chap. 2) outlined how collaborative problem solving consisted of cognitive and social domains. The cognitive domain consisted of skills in task regulation and knowledge building. The social domain could be explored through a person's participation, perspective-taking and social regulation. The hierarchy of the problem solving skills has been defined by Griffin (2014).

Problem Solving

In addition to collaborative problem solving and learning through digital networks, the project team in London identified non-collaborative or individual problem solving as a third area to be explored. Research into this area began as a Piagetian approach to developmental growth and problem solving but, due to time and budget constraints, it was discontinued in mid-2010. During the construction of collaborative

Table 1.1 The expert panels

Problem solving		
Beno Csapo	University of Szeged, Hungary	Panel Lead: Problem solving
Philip Adey	King's College, London	Problem solving
Jarkko Hautamaki	University of Helsinki	Problem solving
Terezinha Nunes	University of Oxford, UK	Problem solving
Patrick Griffin	University of Melbourne	Measurement
Collaborative problem solving		
Friedrich Hesse	Knowledge Media Research Center (KMRC) University of Tubingen	Panel Lead: Collaborative problem solving
Eckhard Klieme	German Institute for International Educational Research	Collaborative problem solving
Marlene Scardamalia	University of Toronto	Collaborative problem solving
Kurt Vanlehn	Arizona State University	Collaborative problem solving
Esther Care	University of Melbourne	Measurement
Learning through digital networks		
John Ainley	Australian Council for Education Research	Lead: ICT Literacy
Kathleen Scalise	University of Oregon	ICT Literacy
Peter Pirolli	Palo Alto Research Center	ICT Literacy
Jean-Paul Reeff	German Institute for International Educational Research	ICT Literacy
Mark Wilson	University of Berkeley	Measurement

problem solving tasks at the University of Melbourne, however, extensive use was made of a series of inductive and deductive reasoning tasks in a problem solving study (Griffin et al. 2013).

The members of the three substantive expert panels are listed in Table 1.1.

Once the skillsets were defined and operationalized with the hypothesised developmental progressions and possible assessment strategies, lobbying began by the project executive, task force and advisory board to influence governments, corporations and evaluation agencies. A request was made to the PISA governing board that collaborative problem solving be considered as a national option in 2012. However, the planning cycle for PISA studies had already begun for 2012 and collaborative problem solving could not be included in the PISA 2012 study, even as a national option. Consequently, discussions between the ATC21S executive board members and the OECD focused on the possibility of assessing collaborative problem solving as a core skill in 2015. This measure was adopted. The planning cycle of the International Association for the Evaluation of Education Achievement (IEA) could not be changed so the new, more complex definitions of ICT literacy could not be introduced into their study in 2013. However, the association was fully supportive of the new direction in which the ATC21S project was taking ICT literacy.

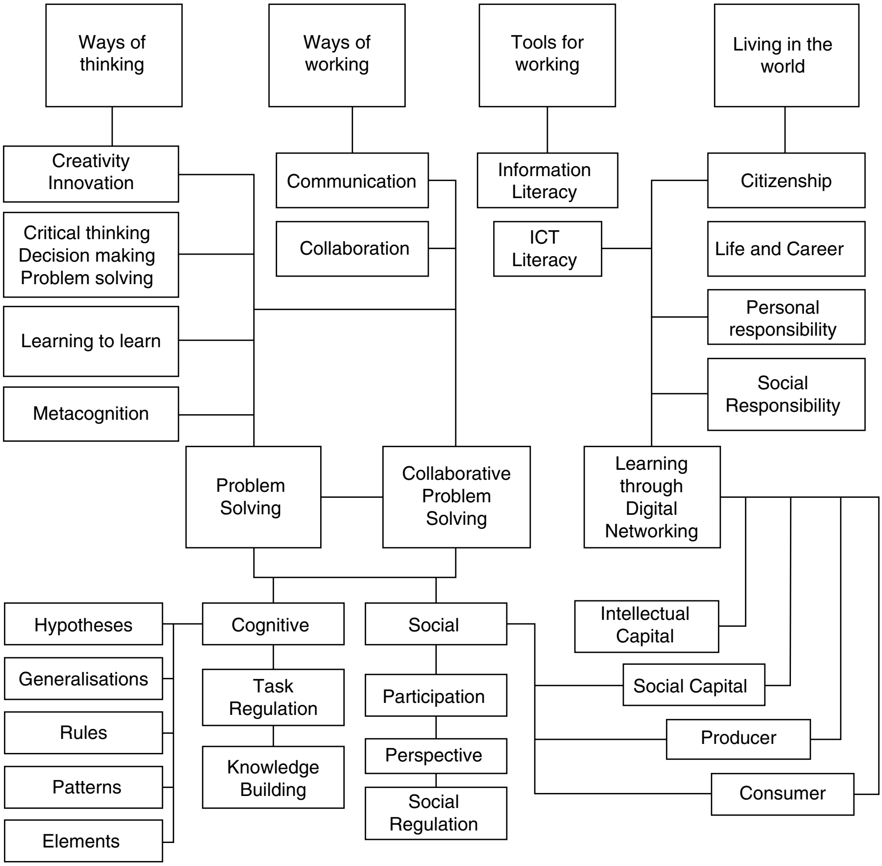


Fig. 1.1 The twenty-first century skills framework

The hypothesised relationships between the skills of problem solving, collaborative problem solving and learning through digital networks are illustrated in Fig. 1.1, which illustrates how the twenty-first century skills and their components are expected to be interrelated and correlated. This was tested during the research phase of the project and results will be reported in ATC21S Volume 3 by Springer. The relationship among these twenty-first century skills and the way in which they build to describe the individual person is illustrated in Fig. 1.2.

Psychometric Panel

The second panelling exercise consisted of a review by psychometric specialists. Their task was to review the assessment tasks and the quality of the data generated by the tasks. Their ultimate goal was to assess the suitability of the data for scaling and empirical analysis. This role became important during the trials of the tasks.

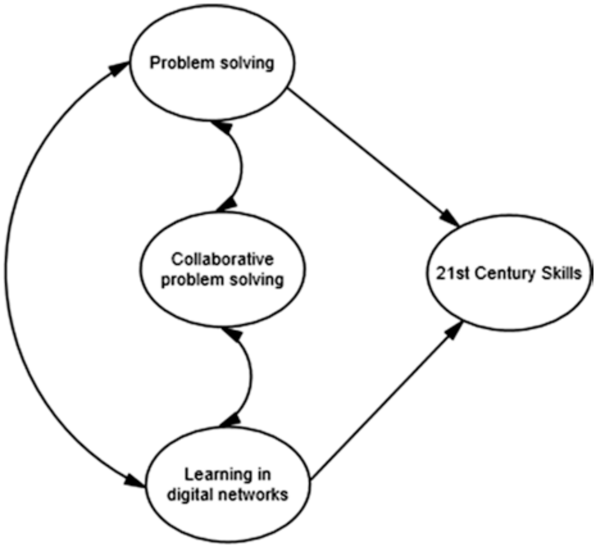


Fig. 1.2 Methodological requirements of assessment

The psychometric panel consisted of the executive director (Patrick Griffin), International Research Coordinator (Esther Care), Ray Adams (University of Melbourne) and Mark Wilson (University of California, Berkley).

Teacher Panels

Panels consisting of teachers were organised by the national project managers within each of the participating countries. Their task was to review the draft materials, oversee the cognitive laboratories and pilot the tasks in order to provide information about infrastructure and administration. The teachers reviewed the tasks and the qualitative data derived from the activities. They expressed their views about the learning and teaching potential of each task. They provided qualitative feedback within country to national project managers, which was then summarised and relayed to the international research coordinator for input to the task development process.

Education Context

In general, the development of the tasks was governed by the requirements for good education measurement. This meant that they needed to be related to a well-defined construct. They needed to provide a sense of a developmental progression and this in turn needed to be based on research and theoretical evidence. The draft developmental progression was supported by empirical evidence from trials pilot studies, including cognitive laboratories. Following Hattie (2011) the evidence needed to enable the students' thinking process to be visible. Some of the challenges included the necessity of learning more about developmental progressions using data on processes rather than products, solutions and outcomes. The process used by the student was considered to be as important, if not more important, than the problem solution. Developers were instructed to distinguish the context in which the task was set from the cognitive construct that was being assessed. This is particularly important in generic kinds of capabilities because they are typically embedded within content that is generally familiar to teachers and students as part of the existing curriculum. There was a need to develop ways in which teachers could be assisted to use data from these assessments in a formative mode to improve both teaching and learning. Bearing in mind that the title of the project was the Assessment and *Teaching* of 21st Century Skills, the idea of formative assessment needed to be intrinsic to the project.

Technical issues challenged team members at the directorate, the developers of the tasks, and the national project managers. There needed to be a way of improving precision and efficiency in data collection for large-scale assessment, as well as providing feedback to teachers and students and systems of education, while allowing the data to be aggregated for purposes of policy consideration. New forms of data collection needed to be devised, and methods of analysing those new forms of data needed to be identified and tested.

Broad issues needed to be examined, such as the migration of classroom-based student interactive and teacher activities to technology-based assessment context. In many cases it was necessary to start with an understanding of the current capacity of the technology and then expand the possibilities that were available to the project team. Viewed in another way, the process began with a statement of an assessment need and then sought a technological solution. A classic example is the notion of collaborative work in the classroom. It was necessary to assess collaboration by observing or measuring participation, involvement or engagement in a task by a class of 30 students working together, discussing the problem and talking to each other. All this activity around collaboration and participation was to be captured digitally, in the background of the task administration.

If a test is to be used in a high stakes environment, security issues become predominant. With ATC21S, ways of ensuring the test security, maintaining data security and ethical use of data, as well as scoring and protecting the identity of test-takers, became issues of administration in terms of functional aspects of the portal design for access to the tasks and the support materials. The project team needed to develop a model and a interpretation procedure that enabled coding of task procedures and solution in addition to identifying ways of documenting and modelling the students' use of problem solving and collaboration procedures. Access to and control of data remains a serious issue.

Classroom environments play an important role in the assessment methods for twenty-first century collaborative work. Schools have been recognised as knowledge-building organisations, particularly in the development of assessments of learning through digital networks. It was accepted by the project teams that students can and do develop their own skills and build knowledge. The challenge was to identify and record those procedures so that they could be coded and interpreted in ways that were useful for teaching and learning. In keeping with a view of twenty-first century education outlined in Griffin et al. (2012a, b), twenty-first century education needs to be knowledge-centred, so that students can demonstrate deep knowledge and understanding as key to developing expertise, learner-centred with students actively engaged, community-centred so that knowledge building is collaborative, and assessment-centred so that progress can be monitored. It was clear to the project team that progress was dependent on the availability of a road map in the form of a developmental progression. Hence there were the early instructions given to the substantive expert panels: "describe at least three levels of development to ensure that measures can have order, magnitude and direction". Once the empirical data was obtained and calibrated (Griffin et al. 2015; Chap. 7) the measures would have the fourth property (Wright and Masters 1982) – a repetitive unit of measurement.

The project team required of the commissioned task developers that they take advantage of the evolution of the internet, using Web 2.0 to allow production and interaction among the participants, and to anticipate the evolution of Web 3.0, in which the internet would gather intelligence about the participants and adjust the interaction process accordingly.

Policy and Expectation

It was necessary to be aware of the emerging policy frameworks (Darling-Hammond 2012). Many new assessment approaches have been developed in the past but there is a history of failure among new innovations. It is often difficult to take a new approach, particularly in the field of assessment, and scale it up to international or global level. One of the goals of the project was to make sure that the assessments would scale. It was necessary for the developers and the project team to be aware of policy frameworks that could be used to facilitate successful wide-scale adoption of the assessment. Adamson and Darling-Hammond (2015; Chap. 15) were commissioned to identify the kinds of strategies that would help to integrate school-based and large-scale assessment and consider how these might scale to become national or international approaches.

It was expected that the project would lead to advances in technology-based assessment. There were expectations that there would be a focus on social and other non-cognitive skill development. There was an expectation that there would be cross-national collaboration in the development process. Furthermore it was expected that the public-private partnership process involving industry, government and education systems would serve as a model for future similar developments.

From the beginning, the sponsors of the project made it clear that they were not pursuing this project set of goals for financial gain. They did not envisage commercializing the products at the end of the project and repeatedly reassured the education community that commercialization was not their goal. It was expected that all materials and procedures developed in the project would be made available in the public domain under creative commons attribution. In the period 2009–2011 the white papers (Griffin et al. 2012b) would define the skills, review previous work and identify issues for research and development. This review would stimulate new forms of assessment strategies that could be validated in pilot studies and trials but would not involve or include the development of specific traditional format. Pilot studies and trials were conducted during that period in the participating countries.

In the period 2011–2012 the goal was to link ATC21S to the PISA data collection, to influence national assessments, in the participating countries at least, and link to other cross-national studies.

From 2012 and 2014 onwards the project sought to influence classroom teaching and learning through changes in assessment practices and to have these trialed and studied in several countries. At the time of writing this volume, plans are in place for this to occur during 2014 in Australia, the United States, Singapore, Costa Rica and Finland. Expressions of interest have also been made by governments or academic research organizations from South Korea, China, Japan, India, Russia and Sweden.

Research and Development

The research and development program set out to construct assessments of critical thinking, problem solving and collaboration within a digital environment. The project also had to establish new ways of assessing digital literacy and the ways in which people would learn through digital networks. The assessment tasks were to be suitable for students in the range of 11–15 years of age. This would cover the lower secondary or middle school. It needed to encompass both individual and collaborative skills. There was a need to include simulations in dynamic tasks which change the problem as the students work through them. The tasks and their solutions needed to be technology-based. There also needed to be some opportunity for face-to-face collaboration in the classroom when students were learning to develop the skills. Hence, teaching classroom support materials were needed as a by-product of the assessment program.

Formulating Hypotheses About Progressions

Wilson (2009) argues that the first step in constructing assessments is to hypothesise a developmental progression that underpins the construct to be measured. The substantive expert panels were given three instructions.

1. Describe the construct under investigation. Identify sub-strands and contextual factors that affect the development of skills in these sub-strands. Formulate a

hypothesised developmental progression that may be used by task and assessment specialists to define appropriate materials for assessment purposes.

2. Describe at least three levels of performance that show the difference between those who possess a great deal of this particular construct and those with very little. (This instruction was to ensure that each of the hypotheses and the progressions had order, direction and magnitude possessed by people. It was also necessary to be able to distinguish between those who possess a great deal of the construct and those with very little amounts of the construct.)
3. Draft ideas about the way in which these particular skills can be demonstrated. That is, describe activities and tasks that would elicit the kinds of behaviour described in the hypothesised developmental progressions.

The emphasis was on formative use of assessment data to improve teaching and learning in the areas defined as twenty-first century skills. The most important outcome may well have been the goal to link the assessment information to direct and explicit teaching and intervention activities. Much of the research being undertaken from 2014 onwards will address these issues.

The Commissioned Task Developers

The project team consistently emphasised that the goal was to introduce innovative, interactive and rich forms of assessment. The assessments would need to be able to monitor individual student growth and development from lower-order skills to higher-order skills within the twenty-first century skills set defined in Volume 1 (Griffin et al. 2012b). The skills underpinning the assessment tasks were expected to be teachable, measurable and applicable to large-scale implementation, and evidence of learning and skills was to be elicited through classroom tasks. The tasks should involve elements of ambiguity and the use of multiple resources and should engage students in ways that ensured they were dependent on one another for successful resolution.

The assessments should enable teachers to adapt the tasks to local context and still measure the developing proficiency in twenty-first century skills. They should enable teachers to interact and teach while tasks are being completed by the students. The opportunity to teach and to promote learning while tasks are being undertaken is an important departure from many high-stakes, high-pressure forms of assessment. For this reason there was an emphasis, in the instruction to task developers, to focus learnable and teachable skills rather than personal attributes. An important aspect of the task construction was that the assessment data should allow formative decisions to be made by the teacher to help improve both teaching and learning. Hence the task developers were required to develop prototype tasks that were usable for a variety of purposes and functions in education, ranging from the individual student or collaborative group to the classroom, the school or the education system. It was necessary to be able to track progress across levels defined within the developmental progressions. There needed to be a process of automatic scoring of all tasks. It was not acceptable in ATC21S that the teacher be required to undertake the direct assessment observation and scoring of the student involvement. Developers were instructed that background student activity data needed to be monitored so that it could be coded, scored, calibrated and interpreted in a way that was independent of the teacher. The teacher's main involvement would be to administer the classroom arrangements for the assessment and to interpret reports in terms of the type of teaching instruction and intervention best suited to the student work. This foreground feedback to students and teachers for purposes of formative intervention would involve a mix of task types and a range of data would be provided to the teacher in the form of skills the students were ready to learn (not a score or a percentage or a grade).

So the developers needed to follow the steps of defining a proof of concept, providing draft copies of the materials to the project team for dissemination for panelling. The task was to enable the national project managers to explore and evaluate the ATC21S project materials in the context of local curriculum. Advice was passed from the national project managers to the International Research Coordinator to ensure that the tasks were sufficiently robust to be applied in each of the participating countries. This also helped to ensure that the opportunity existed for the new assessment prototype tasks to become models for future assessment procedures and that they would link to local and prevailing curricula in each of the countries. It was also important to gather evidence that the tasks would enable teacher intervention during the process of the completion of the tasks to promote and enhance understanding.

National Project Manager (NPM) Role

The National Project Managers (NPM) facilitated feedback from teachers concerning task concepts; data from cognitive laboratories with students; qualitative task response data from students in pilot and in trial stages; and iterative feedback from teachers on the progressions. They were also required to establish and maintain connections with schools; maintain working relationships with teachers; familiarise themselves and teachers with a variety of pedagogical approaches pertinent to twenty-first century skills development among the students; develop a working knowledge of cognitive processing models; monitor and supervise classroom data collection during cognitive laboratories, pilots and trials; and have an understanding of a developmental approach to learning and assessment. The project was largely dependent on national project managers implementing and coordinating the procedures that were asked of them in order to deliver the outcomes required. On several occasions national project managers were expected to attend workshop sessions in their own country and elsewhere, in order to standardise procedures for the project across the participating countries. Their major role was to liaise with the international research coordinator to facilitate procedures and standardise practices across countries. The roles of the NPMs are outlined in Table 1.2.

Table 1.2 Roles of the national project managers

Task concepts
Estimated number of task concepts: 6 at each of four age/grade levels
Estimated number of teachers: 3 at each of four age/grade levels
Contact schools/teachers for participation
Conduct overview session with teachers
Conduct year-level sessions with teachers
Record and aggregate data and return to teachers for verification
Synthesise data and send to directorate
Cognitive laboratories
Estimated number of tasks: 4 within each skills area at each of four age/grade levels (12 at each level)
Estimated number of students: 6 at each of four age/grade levels
Contact schools/teachers for participation
Conduct year-level/task type training sessions with teachers
Observe selected cognitive laboratory sessions
Collect data from teachers and record
Synthesise data and send to Directorate
Pilot bundled tasks
Estimated number of tasks: 12 at each level
Estimated number of classes: 1 at each of four age/grade levels
Contact schools/teachers for participation
Conduct task administration training sessions with teachers
Observe selected classes
Collect and record data
Check and send data to Directorate
Trial task sets
Estimated number of tasks: 12 at each level
Estimated number of classes: 1 at each of four age/grade levels
Estimated number of schools: 20
Contact schools/teachers for participation
Conduct task administration training sessions with teachers within schools
Observe selected classes
Collect, record and clean data
Check and send data to Directorate

The Development Process

Figure 1.3 illustrates the complexity of the development process. Three organisations were commissioned to develop the tasks. The Education Assessment Research Centre at the University of California at Berkeley (BEAR) was commissioned to develop the tasks defining learning through digital networks. World-Class Arena in

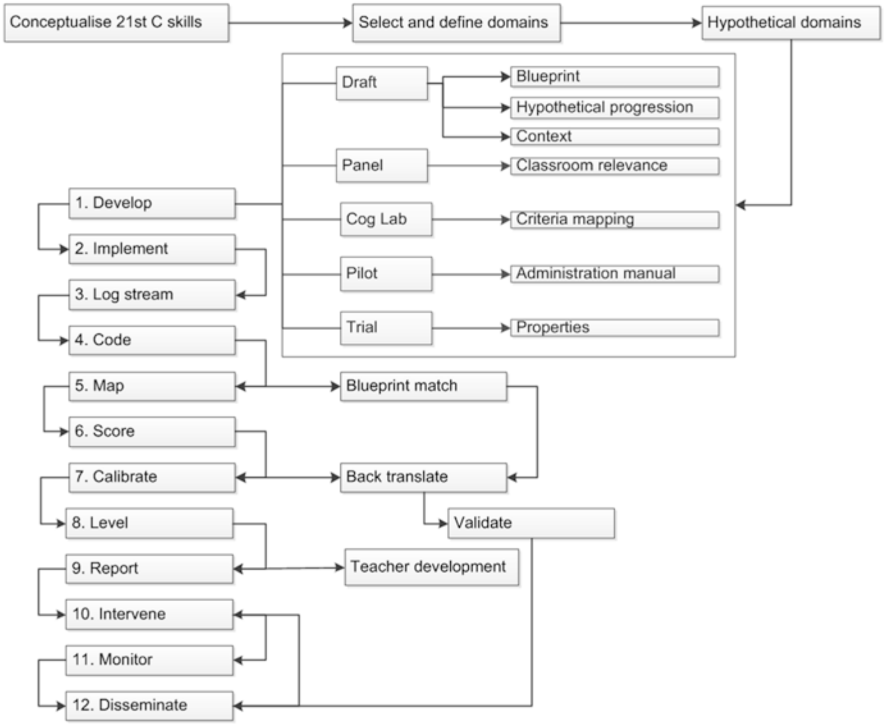


Fig. 1.3 The ATC21S process

the United Kingdom was commissioned to develop collaborative problem solving tasks based on mathematics and science curriculum topics. The University of Melbourne's Assessment Research Centre was commissioned to develop collaborative problem solving tasks based on inductive and deductive reasoning and independent of curriculum content.

Task developers were provided with a theoretical framework for each of the skill areas. Their first task was to suggest a scenario which might help sample the skillset described in the hypothetical progressions. They were to identify specific subskills of interest while working with the expert teams. This enabled them to identify contextual elements relevant to the completion of the tasks which would generate sufficient data to enable the interpretation of student performance in terms of progress along the developmental continuum.

As with most test development, the target student group for each of the complex tasks had to be identified so that appropriate trials of the task could be undertaken with appropriate target and off-target groups of students.

Once the prototype tasks were received, the project directorate implemented the process of the research and development program underpinning the measurement of twenty-first century skills.

Drafting

Given the instructions provided to the developers, it was necessary to ensure that the learning progressions identified for the twenty-first century skills were embedded in the learning environment and in the assessment tasks. The tasks themselves needed to enable scaffolding by the teachers of the student learning process along the developmental progressions and to enable the identification of indicative behaviours that could be used for interpreting the student performance. These data would be necessary for the teacher to interpret the student performance and growth in the skills. But it was also important that the teacher was able to identify the kinds of strategies and intervention resources that would be required to improve both teaching and learning of the twenty-first century skills. So the assessment and teaching strategies involved in ATC21S needed to incorporate systems that supported students and teachers in tagging and analysing their skill development and analysing the key components of their work (e.g., problem solving episodes, evidence gathered, etc.). It was also critical that questionnaires and rating scales would be available to help teachers rate the work within the environment independently and compare results with computer-generated reports.

Panelling

Panelling occurred on three levels. The first was the use of teachers' knowledge of content and curriculum relevance to indicate whether a task was usable, whether the skills could be learned, whether the skills could be taught, and whether the task would identify differences between students who work at a high level and those who work at a low level. In other words, the aim was to discover whether the task would discriminate between those who could and those who could not collaborate either in ICT or in problem solving. Second, teachers were asked to provide some advice about the possible teaching implications and applications of the material in the classroom. Third, the teachers provided evidence of the face validity of the developmental progressions. They were asked to draw on classroom evidence to clarify and modify the descriptions in the progressions.

Advice was also sought regarding the classroom management that would be required to undertake large-scale assessment based on technology-dependent materials. It was important that the project facilitated interaction between practitioners and researchers. The project would introduce new forms of assessment and it was important that teachers felt comfortable with these forms of assessment, not threatened by them, and that the assessments were considered relevant to their classroom environment. National project managers were asked to visit schools, observe what took place in the classroom during the assessments and study the interaction patterns between teacher and student to identify indicators of knowledge development. The process afforded the opportunity for national project managers and field workers to examine multiple literacies (textual, graphical, video, content-specific contextual materials). Their advice was sought about how to enhance student engagement and record indicators of engagement, either through direct observation or through technical supplement. Teachers were asked for their advice about how formative feedback might be provided and referenced to the development of the learning progressions that supported teaching and learning. The end result of this was the collection of data that could be 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.

Cognitive Laboratories

The term *cognitive laboratory* is used to describe the process of observing an individual working through a task and providing metacognitive data during the process. In order for this to occur, the NPM or nominee (task administrator/observer) needed to have a clear understanding of the cognitive processes assumed to underpin the task and to possess communication skills appropriate for interacting with the individual taking the task. The cognitive laboratory process was designed to collect information about the way students went about engaging with the task, to identify any functional issues that needed to be addressed in fine-tuning the task and to identify teaching implications. It also enabled observers to monitor the kinds of cognitive activity that the students were exhibiting and these records could then be matched to the log stream data which was automatically captured as part of the digital assessment system, for interpretive purposes. Individual students who participated were selected from the lower, middle and upper third of the class respectively in terms of general academic progress. This was done to help ensure the required range of metacognitive information was obtained. The classification of the students was the responsibility of the classroom teacher.

The cognitive laboratory data, when provided to the task developers, enabled them to note the kinds of coding that might be used to record the students' actions and reactions and eventually to be able to score student performances. Workshops were conducted for NPMs to ensure that standard procedures were followed and data summaries were reported to the project team via the International Research Coordinator. The developers used the data and project team advice to adjust, abandon or refine tasks and coding protocols.

Piloting

The pilot study process was designed to ensure that the administration, completion and scoring of task functions could be undertaken effectively and efficiently. For this stage, the tasks were presented to students in a penultimate format similar to that in which the final trial would be implemented. The administration of the tasks at this stage was semi-automated, and the NPM (or nominee) was responsible for ensuring that teachers could observe the efficiency and ease of the process, and provide feedback. Ideally, a full class (up to 30 students) from each year level would participate in this process. The focus was on the classroom administration procedures, the technology infrastructure required, the times that the tasks could be expected to take and the observed relative difficulty of the tasks for the students. The documentation that was possible as a result of the pilots essentially formed the basis of the user's manual for assessment task administration.

Trial

This process was designed to provide sufficient data upon which to establish empirically-based scales that had the capacity to indicate students' location and progress on the developmental continua. The administration of the tasks for the trials was semi-automated, allowing for, but not necessarily requiring, teacher input apart from their fulfilment of a supervisory role. The NPM had responsibility for recruiting schools to participate and ensuring that the school infrastructure and facilities could support the trial.

The major purpose of the trials was to establish the psychometric properties of the tasks. It also enabled the interpretation of the student performances in terms of their location on a developmental continuum. Additionally, it enabled reports to be developed for teachers and students. Each of these elements is discussed in the text that follows. In order to conduct trials the delivery platform had to be built. The structure of the platform had to incorporate the capacity for collaborative interaction between students via the Internet. Given the nature of the tasks it was also necessary that a game server be a part of the platform. A detailed explanation of the platform is provided in Awwal et al. (2015; Chap. 5).

The Data Capture System

The system was established with an architecture that allowed students to access the tasks in pairs (dyads), for both problem solving and digital networks. Teachers were able to log on after a suitable time and request student, class and school reports. The request triggers a real time scoring and calibration process. While the students are completing the tasks a log stream file records information about the event. This includes student and partner identity, gender, country; and task identification and specific page within task. Each action and chat episode that each student undertakes. Recorded is whether the student is Student A or B of the dyad, the action taken by the student, start and finish time of actions, chat, progress within a task, and a description of the action or the chat. Within ATC21S there is no interpretation of the chat itself but some text and numbers are searched for, and inferences of process are drawn from action and chat sequence. Activity is also time-stamped for monitoring purposes. In the development phase, these logged data analytics were then coded and mapped onto the task blueprint in order to interpret the activity of each student in reference to the conceptual framework.

Element	Indicator	Low 0	Middle 1	High 2
Task Regulation				
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
Set goals	Sets a clear goal for a task	Sets general goal such as task completion	Sets goals for subtasks	Sets goals that recognise relationships between subtasks
Resource manag[illegible]ent	Manages resources or people [illegible] sk	Uses /Identifies resource[illegible] direct[illegible]	Allocates people or resources to a task	Suggests that people or [illegible]s be used

Fig. 1.4 Section of the blueprint for the collaborative problem solving tasks (See Hesse et al. 2015)

Mapping to the Blueprint

The detailed blueprint for the collaborative problem solving tasks is described in Hesse et al. (2015; Chap. 2). A section of it is illustrated in Fig. 1.4 to demonstrate its organisation – it included a description of each of the 21 elements, in indicative ("Indicator") terms, followed by nutshell statements of how each element might be seen ordered from low to high sophistication. These statements were used by the developers to prompt the creation of processes or actions that could be enacted by a student dyad to solve a problem, and that would generate records of these actions that could be coded against the blueprint.

Coding

The approach to coding the log stream data is provided in detail in Adams et al. (2015; Chap. 6) in this volume. Both sequences of actions are inferred to represent particular skills, as well as direct actions. For example, an inference about an element of social behaviour was taken to be the presence of chat before an action took place. In the Laughing Clowns task (Care et al. 2015; Chap. 4), in which the two students are presented with 12 balls to test the similarity of how their two clowns function, this was assigned the code U2L001 (Table 1.3), which identified the task (U2) the action local, or specific, to this task (L) and the indicator number (001). With the student identified who performed the action included, a letter (A or B) was added to

Table 1.3 Example of scoring rules applied to the coded data from the log stream

Indicator ID	Element	Rule	Description	Scoring	Coding
U2G26A (or B)	15	Asking partner questions	Presence of "what", "how", "who" "where" "why" "when" or "?"	Presence/ absence	1=1, else 0
U2L001A (or B)	2	Communicating before acting	Presence of chat before any moves/ actions	Presence/ absence	1=1, else 0
U2L004A (or B)	13	Undertaking activity (understanding implicit instructions, regardless of partner role)	All positions have been covered (provided player has at least 3 balls)	Presence/ absence	3 positions covered = 1, else 0, IF dropshute<3 then missing
U2L006A (or B)	13	Systematic behaviour (trial of all combinations of resource, sequential)	Sequential placement of balls (6 combinations: LMRLMR, RMLRML, LMRRML, RMLLMR, LLMMRR, RRMMLL)	Presence/ absence	1=1, else 0

the code. An example of coding of an element of cognitive behaviour from a direct action lies in U2L004A, where the same Clowns task (U2) is referred to, a local indicator (L), the indicator number (004), and Student A (A). This indicator codes the testing of all positions for the route taken by balls placed in the Clown's mouths by Student (A), at a point when that Student A has access to at least three of the 12 balls. This action implies a relatively systematic approach to the exploration of the problem.

Conversion and Scoring

Once the codes were allocated, the data could be converted to a score file for use in data analysis and task calibration. To achieve this, scoring algorithms were developed. The purpose was to score every action and chat separately and in combination. The coded indicators became variables or items in the data assigned to each student. Each indicator was recorded as present or absent. Frequency was then interpreted as a proxy measure of difficulty. With these linked to the difficulty levels in the blueprint, further interpretation is possible, taking into account partial credit analyses. Table 1.3 illustrates how scoring rules were applied to the Clowns task. Additional information is provided in Adams et al. (2015; Chap. 6).

Calibrating

Once the data was scored, a calibration analysis was undertaken. This enabled the ability of each student and the relative difficulty of the indicators to be estimated. Once this was achieved, it enabled the estimation of student ability on the basis of a subset of indicators from a selection of CPS tasks rather than requiring the student to complete all tasks in order to obtain the ability estimate (see Adams et al. 2015). Calibration of the collaborative problem solving tasks was undertaken to investigate the five strands of the construct. Figure 1.5 presents the item or indicator difficulty and the fit to the Rasch model of the Clowns task (Griffin et al. 2015; Chap. 7). The excellence of the fit is evident (Table 1.4).

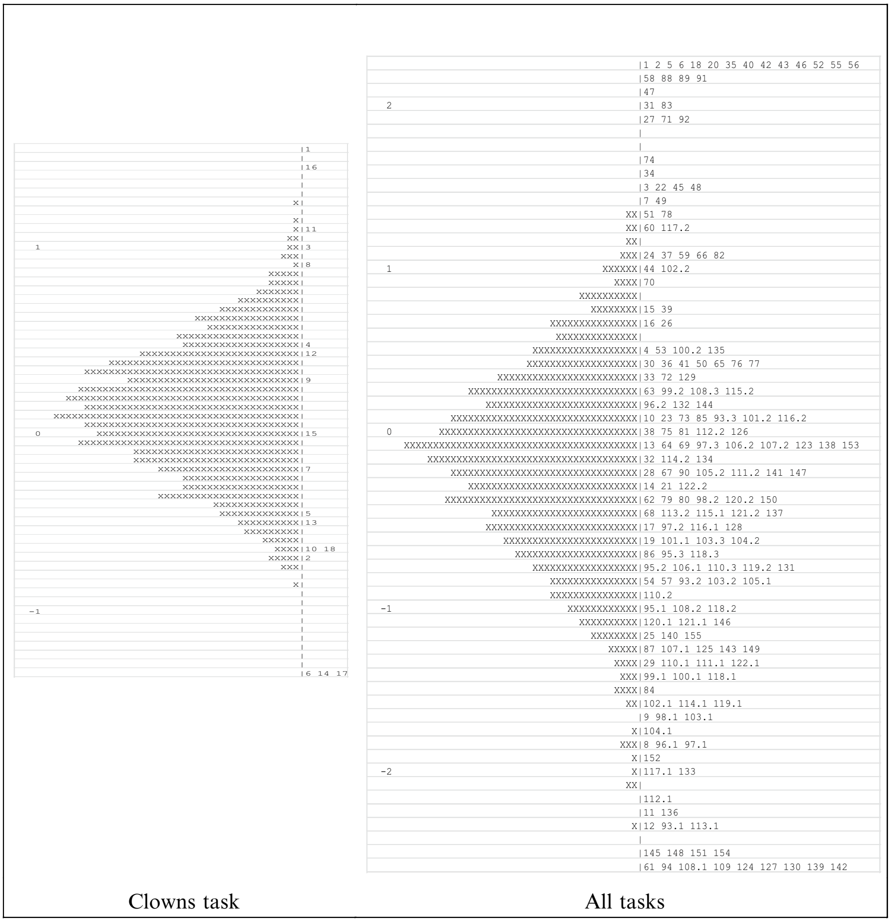


Fig. 1.5 Variable maps for the Clowns task and the concurrent calibration of all tasks

Table 1.4 Illustrative item difficulty and fit to the Rasch model of the Clowns task (See Griffin et al. 2015)

Variables				Unweighted fit				Weighted fit			
Item		Estimate	Error	MNSQ	Confidence interval		T	MNSQ	Confidence interval		T
1	1	3.106	0.046	1.13	0.91	1.09	2.8	1.02	0.75	1.25	0.2
2	2	−0.686	0.04	0.97	0.91	1.09	−0.5	0.98	0.95	1.05	−0.7
3	3	1.01	0.04	1	0.91	1.09	−0.1	1	0.94	1.06	0
4	4	0.454	0.039	1	0.91	1.09	−0.1	1	0.97	1.03	−0.2
5	5	−0.435	0.039	1	0.91	1.09	0	1	0.96	1.04	−0.1
6	6	−1.409	0.042	0.98	0.91	1.09	−0.5	0.99	0.9	1.1	−0.2
7	7	−0.218	0.039	1.01	0.91	1.09	0.3	1.01	0.97	1.03	0.9
8	[illegible]	0.895	0.039	0.98	0.91	1.09	−0.4	0.99	0.95	1.05	−0.5
9		0.267	0.039	1.06	[illegible]	1.09	1.3	1.06	0.98	1.02	4.5
10		−0.657	[illegible]	1		[illegible]	0		0.95	[illegible]	0
11		[illegible]		[illegible]		[illegible]					−0.5
12											2.8
											[illegible]

Each of the tasks was separately calibrated and individual variable maps are presented in Griffin et al. (2015; Chap. 7). Figure 1.5 represents the Clowns variable map on the left hand side and all tasks on the right. It is presented here to illustrate the point that each task can represent the construct to different degrees. Some tasks have been designed to indicate slightly higher level skills across some elements, while others have been designed to focus more on particular elements. This design approach provides the facility for the grouping or "bundling" of tasks for students to engage with in order to provide a robust measurement of their skills. The map shows the distribution of students in terms of their ability on the left hand side of each map, and the numbers on the right hand side represent the item numbers. As can be seen the Clowns task samples within one part of the ability range. When a task such as this is grouped or bundled together with other items, the robustness of the measurement is strengthened. Once each task was calibrated, all tasks were concurrently calibrated to check that they all mapped onto the same underlying measure and construct. Concurrent equating was used to ensure that it was a matter of indifference which set of tasks was undertaken. The concurrent calibration variable map is shown on the right in Fig. 1.5.

The examples reported here are based on data collected during the ATC21S trials. These data were collected by participating countries based on convenience samples, and in no way are presumed to be representative of the populations of these countries.

The concurrent equating was undertaken at three levels – for one unified construct, for the social and cognitive components, and for the five strands (participation, perspective and social regulation; and task regulation and knowledge building). Figure 1.6 illustrates the complexity of the construct, with the clusters of indicators on the right of the figure used to interpret the underlying dimensions.

The five collaborative problem solving strands are correlated, but with each contributing unique variance (Table 1.5). Their degree of correlation provides important information regarding the possible transfer of skills from one dimension to another. Also, once the student can be mapped onto the developmental progressions for the five strands, teachers can use the separate strand descriptions to plan instruction and intervention.

Each task differs in terms of which elements of the blueprint it maps onto and at what difficulty levels. In addition, students can take different approaches or paths to

```
-------------------------------------------------------------------------------------------
                |         |         |         |         |                                 |
   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            |         |         |         |         |                                 |
                |         |         |         |         |                                 |
                |         |         |         |         |                                 |
                |         |         |         |         |                                 |
                |         |         |         |         |                                 |
                |         |         |         |         |                                 |
===========================================================================================
```

Fig. 1.6 Five dimensional map of the collaborative problem solving construct. Each '`X`' represents 45.8 cases

Table 1.5 Correlations between the estimates of student ability across the five strands

Strand	1	2	3	4	5
Strand 1					
Strand 2	0.565				
Strand 3	0.850	0.332			
Strand 4	0.781	0.549	0.748		
Strand 5	0.703	0.683	0.482	0.693	

Table 1.6 Raw score and parameter estimates for student (see Griffin et al. 2015; Chap. 7)

Person ID	Raw score (social)	Maximum possible	Raw score (cognitive)	Maximum possible	θ_{social}	$\theta_{cognitive}$
1	41	47	32	57	1.028342	–0.26011
2	33	47	38	58	0.039957	0.097414
3	35	53	40	60	0.088804	0.19713
4	26	46	27	57	–0.54873	–0.62936
5	35	46	39	58	0.267853	0.262837
6	33	47	32	58	0.039961	–0.38134
7	32	52	4[illegible]	65	–0.12961	0.786059

the problem solution process. For example, the maximum score in Table 1.6 represents the maximum number of relevant indicators that can be demonstrated for any specific task or collection of tasks. There are different numbers of indicators demonstrated by different students depending on their approach to the task. Hence the ability estimate is based on which indicators are demonstrated by the student, and the p = maximum possible indicators given the set of tasks attempted and role adopted by the student (Student A or B). Table 1.6 presents an extract from a set of raw scores for both social and cognitive strands and the Rasch model estimates of the student ability. It is clear that different students had different possible maximum numbers of indicators to demonstrate and varied in the actual numbers of indicators demonstrated.

Estimating Student Ability Parameters and Determining the Levels

Once the indicators were mapped onto the elements and their relative difficulty estimated, the student ability could be estimated using the difficulty parameters. This was achieved using the following algorithm.

$$\theta_n^{(t+1)} = \theta_n^{(t)} + \frac{r_n - \sum_{i \in \Omega_n} e_{ni}^{(t)}}{\sum_{i \in \Omega_n} v_{ni}^{(t)}}$$

and stop when $|\theta_n^{(t+1)} - \theta_n^{(t)}| < .001$.

Once the estimates of student ability are obtained it is possible to map each student onto a level within a developmental progression as described by Griffin et al. (2015; Chap. 7). We take as an example the knowledge building strand of the cognitive domain of collaborative problem solving. Some items drawn from different tasks shown in Table 1.7 link the estimate of difficulty in logits, derived from the Rasch calibration, to the raw score for each task. The element against which the item is mapped, the indicator code and description are provided, as well as the scoring rule for assigning a score to the indicator. The "torn" line marks the threshold between

Table 1.7 Interpretation of student performance in the knowledge building thread based on illustrative items (Note. Items have been selected from different levels of difficulty as illustrated by the parameter estimates)

δ_i	Element	Indicator code	Indicator description	Scoring rule	Score
4.734	21	612A	Problem solved [correctness, mid, subtask, can answer independently]	Answers correctly	Presence or absence
4.028	21	714A	Problem solved [correctness, late, subtask, can answer independently]	Answers correctly	Presence or absence
2.501	21	310B	Correctness [problem solution is independent of partner]	Correctness	Presence or absence
1.638	21	425B	Problem solved [correctness, subtask, late]	Correctness of answer	Presence or absence
0.802	18	815A	Problem solved [correctness, late, partial subtask, can answer independently]	Correctness of horizontal cell rules	Presence or absence
0.49	21	310A	Correctness [problem solution is independent of partner]	Correctness	Presence or absence
0.147	21	010B	Problem solved [correctness]	Answers correctly	Presence or absence
–0.918	19	Q15 – self	When we had different information, I asked my partner for information		

adjacent levels of performance. With this information it is possible to compare the student ability estimate obtained from the calibration algorithm above to thresholds, allowing the student to be located on the development progression called '*knowledge building*'. The thresholds for this strand are −1.6, 0.2, 1.1 and 3.0 for levels 0–1, 1–2, 2–3, 3–4 and 4–5 respectively. The theoretical basis for this interpretation and these cut scores is provided by Griffin (2007).

These cut scores and clusters of indicators enable an interpretation of the variable and the definition of the developmental progression as well as identification of the skills a student is ready to learn given that the response probability for the calibration was set at 0.5 (Griffin 2007). The interpretation for knowledge building that this leads to is shown in Table 1.8. The element notation in Table 1.8 refers to the element number from the collaborative problem solving framework followed by its performance criteria within that element (1–3 representing lower to higher sophistication of skill). For example, 11.3 refers to Element 11 "Collects Information" of

Table 1.8 Levels in the developmental progression for knowledge-building

Levels	Element notation	Knowledge building
Level 6	11.3 13.3 19.3 21.2	The student has a good understanding of the problem and can reconstruct and/or reorganise the problem in an attempt to find a new solution path.
Level 5	10.3 12.3 14.3 15.2 19.2	The student can identify cause and effect and use suitable strategies to gain a correct path solution for both simple and complex tasks. The student can modify and adapt their original hypotheses, in light of new information, testing alternatives hypotheses and altering their course of thinking.
Level 4	10.2 13.2 14.2 17.3 18.3 21.1	The student can identify connections and patterns between multiple pieces of information. The student can successfully complete subtasks and simpler tasks.
Level 3	11.2 15.1 18.2	The student begins to connect pieces of information together.
Level 2	10.1 11.1 12.2 14.1 17.2 19.1	The student tests their hypotheses based on the information they have. They identify possible cause and effect of actions and repeats attempts in order to gain more information about an actions outcome.
Level 1	12.1 13.1 17.1 18.1	The student continually attempts the task with the same approach with little evidence of understanding the consequences of actions taken. The student focuses on each piece of information individually, only following the specific instructions provided.

the collaborative problem solving framework, and the highest performance criterion coded at 3. Table 1.8 clarifies the reality that not all elements are to be found at the same overall level of difficulty; the lowest level of skill in one element might coincide with the highest level of skill in another element.

Reports

This stage of the research led to the formation of the reporting module. A project decision was made not to report a score but to illustrate to student and teachers what the student is most ready to learn (Griffin 2007). A summary statement for each level was entered into the reporting module. If the student ability estimate is derived as outlined above, a series of reports can be generated.

Once the data were entered, the reporting module can be initiated by a teacher. A series of reports can then be produced. The *Learning Readiness Reports*, *Class Reports* and *Student Profile Reports* are illustrated and explained in Woods et al. (2015; Chap. 14). The reports (Fig. 1.7) provide both individual and group information, each with a focus on skills location and skills progression. In each case, the teacher has access to information that provides an estimate of the current functioning of a student, and an outline of the skills that the student is most likely to be able to develop next. Teachers are able to use this information to plan how and what to present to facilitate student learning.

Teaching Twenty-First Century Skills

The system is designed so that once the students have undertaken the tasks the teacher can log onto the Internet portal and request a download of student reports as shown in Fig. 1.7. Once the teacher generates the reports, the most important

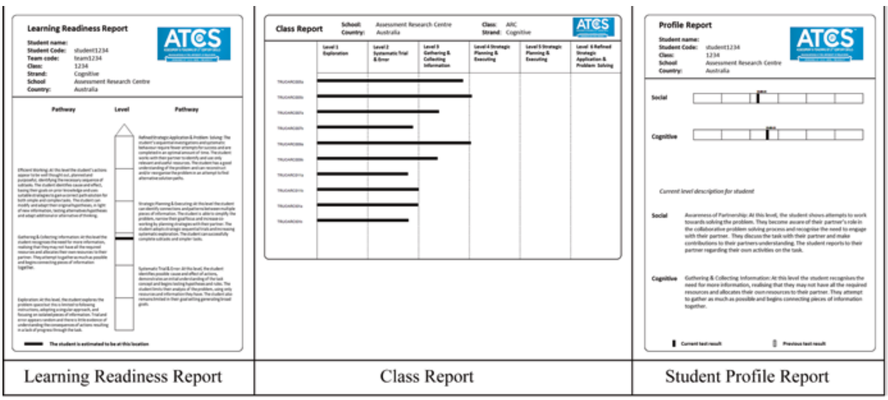


Fig. 1.7 Samples of reports

aspect of ATC21S begins: the teaching and learning of twenty-first century skills. The development of the student skills and the teaching interventions are outlined in a series of professional development modules. The professional development modules are available for teachers online and cover the following:

1. Defining and Assessing Twenty-First Century Skills
2. Using a Developmental Model
3. ATC21S Assessments: Getting Started
4. Interpreting Reports
5. Teaching and Learning Twenty-First Century Skills

Conclusion
The research and development process underlying the ATC21S project followed a typical test development path. The constructs of interest were defined and described. How students might perform at different levels of skill on these was hypothesised, and tasks were designed to sample these performances. The focus was on constructs of interest in the twenty-first century workplace, and therefore in the twenty-first century classroom. The underlying ideology was that the function of assessment was to stimulate change. This means that the constructs of interest were those that could be shown to vary across individuals, and that would be amenable to change. Teachers' understanding of how students might vary in their twenty-first century skills can be informed by their observations of students when engaged in both online and classroom based tasks, and by their analysis of their students' results. Changes in student performance are hypothesised to be activated by teaching. The full process of the development behind creation of assessments of twenty-first century skills has been outlined. The next steps in the research and development process are to identify how teachers will go about teaching these skills.

References

Adams, R., Vista, A., Scoular, C., Awwal, N., Griffin, P., & Care, E. (2015). Automatic coding procedures for collaborative problem solving. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills: Methods and approach* (pp. 115–132). Dordrecht: Springer.

Adamson, F., & Darling-Hammond, L. (2015). Policy pathways for twenty-first century skills. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills: Methods and approach* (pp. 293–310). Dordrecht: Springer.

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

Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., & Rumble, M. (2012). In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills*. Dordrecht: Springer.

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

Darling-Hammond, L. (2012). Policy frameworks for new assessments. In *Assessment and teaching of 21st century skills*. Dordrecht: Springer.

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

Griffin, P. (2014). Performance assessment of higher order thinking. *Journal of Applied Measurement, 15*(1), 1–16.

Griffin, P., Care, E., & McGaw, B. (2012a). The changing role of education and schools. In P. Griffin, B. McGaw, & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 1–15). Dordrecht: Springer.

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

Griffin, P., Care, E., Bui, M., & Zoanetti, N. (2013). Development of the assessment design and delivery of collaborative problem solving in the assessment and teaching of 21st century skills project. In E. McKay (Ed.), *ePedagogy in online learning*. Hershey: IGI Global.

Griffin, P., Care, E., & Harding, S. (2015). Task characteristics and calibration. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills: Methods and approach* (pp. 113–178). Dordrecht: Springer.

Hattie, J. (2011). *Visible learning for teachers: Maximising impact on learning*. New York: Routledge.

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* (pp. 37–56). Dordrecht: Springer.

Kozma, R. (2009). Assessing and teaching 21st century skills: A call to action. In F. Schueremann & J. Bjornsson (Eds.), *The transition to computer-based assessment: New approaches to skills assessment and implications for large scale assessment* (pp. 13–23). Brussels: European Communities.

Wilson, M. (2009). Measuring progressions: Assessment structures underlying a learning progression. *Journal of Research in Science Teaching, 46*(6), 716–730.

Wilson, M., & Scalise, K. (2015). Assessment of learning in digital networks. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills: Methods and approach* (pp. 57–81). Dordrecht: Springer.

Woods, K., Mountain, R., & Griffin, P. (2015). Linking developmental progressions to teaching. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills: Methods and approach* (pp. 267–292). Dordrecht: Springer.

Wright, B. D., & Master, G. N. (1982). *Rating scale analysis*. Chicago: NESA.

Part II
Conceptual Framework for Internet Based Collaborative Assessment Tasks

Hesse, Care, Buder, Sassenberg, and Griffin (2015, Chap. 2) and Wilson and Scalise (2015, Chap. 3) present the conceptual frameworks for collaborative problem solving and for learning through digital networks. Each of these frameworks addresses the notion of interaction between students on the Internet in order to solve problems or to formulate and test hypotheses and learn through networking. The chapter on learning through digital networks illustrates how individuals can operate and learn through the social media and it also demonstrates that people working together can shift from being individual consumers and producers of information to collaborative contributors to the development of groups' social capital and intellectual capital. The definition of collaborative problem solving provides a comprehensive analysis of the literature and of the concepts involved in problem solving and collaboration. Both 'Learning through Digital Networks' and 'Collaborative Problem Solving' are shown to be multifaceted, multidimensional, complex and consisting of dimensions that describe the social and cognitive skill development.

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* (pp. 37–56). Dordrecht: Springer.

Wilson, M. & Scalise, K. (2015). Assessment of learning in digital networks. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills: Methods and approach* (pp. 57–81). Dordrecht: Springer.

Chapter 2
A Framework for Teachable Collaborative Problem Solving Skills

Friedrich Hesse, Esther Care, Juergen Buder, Kai Sassenberg, and Patrick Griffin

Abstract In his book "Cognition in the Wild", Hutchins (1995) invites his readers to scan their immediate environment for objects that were not produced through collaborative efforts of several people, and remarks that the only object in his personal environment that passed this test was a small pebble on his desk. In fact, it is remarkable how our daily lives are shaped by collaboration. Whether it is in schools, at the workplace, or in our free time, we are constantly embedded in environments that require us to make use of social skills in order to coordinate with other people. Given the pervasiveness of collaboration in everyday life, it is somewhat surprising that the development of social and collaborative skills is largely regarded as something that will occur naturally and does not require any further facilitation. In fact, groups often fail to make use of their potential (Schulz-Hardt, Brodbeck, Group performance and leadership. In: Hewstone M, Stroebe W, Jonas K (eds) Introduction to social psychology: a European perspective, 4th edn, pp 264–289. Blackwell, Oxford, 2008) and people differ in the extent to which they are capable of collaborating efficiently with others. Therefore, there is a growing awareness that collaborative skills require dedicated teaching efforts (Schoenfeld, Looking toward the 21st century: challenges of educational theory and practice. Edu Res 28:4–14, 1999). Collaborative problem solving has been identified as a particularly promising task that draws upon various social and cognitive skills, and that can be analysed in classroom environments where skills are both measurable and teachable.

This chapter provides a conceptual framework of collaborative problem solving that is informed by findings from fields of research as diverse as cognitive science, education, social psychology and psycholinguistics.

F. Hesse (✉) • J. Buder • K. Sassenberg
Knowledge Media Research Center, Tübingen, Germany
e-mail: fh@iwm-kmrc.de

E. Care • P. Griffin
Assessment Research Centre, Melbourne Graduate School of Education, University of Melbourne, Parkville, VIC, Australia
e-mail: e.care@unimelb.edu.au; p.griffin@unimelb.edu.au

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

Collaborative Problem Solving

Before defining collaborative problem solving, it might be helpful to define the constituents of this term, beginning with "collaboration" and followed by "problem solving".

We define collaboration as the activity of working together towards a common goal. There are a number of elements included in the definition. The first element is **communication**, the exchange of knowledge or opinions to optimise understanding by a recipient. This element is a necessary but not sufficient condition for collaborative problem solving – it requires that communication goes beyond mere exchange. The second element is **cooperation**, which is primarily an agreed division of labour. Cooperation in collaborative problem solving involves nuanced, responsive contributions to planning and problem analysis. An alternative view might regard cooperation simply as a lower order version of collaboration, rather than as a component within it. Our reasons for not adopting this view are provided below. A third element is **responsiveness**, implying active and insightful participation.

From this definition, collaborative problem solving means approaching a problem 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 useful when dealing with problems that are complex.

In the learning sciences there was a major shift in the 1990s to move from "cooperative learning" towards "collaborative learning". While many authors use these terms interchangeably, a key difference was identified by Dillenbourg and colleagues (1996). According to their distinction, cooperation is referred to as an activity which is accomplished through division of labour. In other words, while cooperative learners might coordinate at some points of their activity, they often work in parallel. Many scholars have noted that cooperative learning neither makes full use of a group's potential nor requires the whole set of social skills that people rely on when working together (e.g. Cohen 1994). This led to focus on collaborative learning.

In collaborative learning, learners jointly orchestrate their activities in order to address a particular task or problem. The activities from learners are inextricably intertwined, contributions by learners mutually build upon each other, and one learner's actions might be taken up or completed by another. Only when a task requires collaboration does the full set of social skills come into force. This makes tasks like collaborative problem solving some of the key testbeds for the assessment of 21st century skills.

Problem solving is an activity in which a learner perceives a discrepancy between a current state and a desired goal state, recognises that this discrepancy does not have an obvious or routine solution, and subsequently tries to act upon the given situation in order to achieve that goal state. It is accompanied by a number of mental and behavioural processes that might not necessarily take place in sequential order, but can run in parallel. One approach to conceptualising this notion has been taken by the PISA group in their problem solving framework. First, a problem – that is, a discrepancy between current state and goal state – is identified. Second, a learner makes a mental representation of the problem states and of the steps that allow for a transformation between problem states (typically called a "problem space"). Third, a learner formulates a plan for steps that might enable a move nearer to the goal state. Fourth, the plan is executed. And fifth, the progress towards a problem solution is monitored.

Another, procedural approach implies a solution focus and an awareness of the nature of the problem and the goal states. Griffin (2014) argued that problem solving could be seen as a hierarchical series of steps moving from inductive to deductive thinking. The problem solver first examines the problem space to identify elements of the space. Next they recognise patterns and relationships between the elements, and formulate these into rules. The rules are then generalised and when generalisations are tested for alternative outcomes the problem solver is said to be testing hypotheses. This approach is elucidated in a later section of this chapter.

Based on these definitions and approaches, collaborative problem solving can be defined as a joint activity where dyads or small groups execute a number of steps in order to transform a current state into a desired goal state. The difference between individual and collaborative problem solving is that in collaboration each of these steps is directly observable. Participants need to exchange and share their identification of parts of the problem, their interpretation of the connections between the parts, relationships between action and effect (rules) and the generalisations they propose in search of a solution. The steps towards a collaborative solution may be coordinated through the use of verbal and non-verbal observable signals. Externalisation also has the welcome side effect of making problem solving activities visible and easier to assess.

The stages of individual problem solving apply – though in an altered and more complex fashion – to collaborative problem solving. The implications for the process of involving more than one problem solver in a collaborative context are discussed below.

Collaborative Problem Solving Processes

An idealised depiction of collaborative problem solving could follow a PISA-like sequential process. Collaborative problem solving requires that the collaborating parties recognise a problem and identify which elements of the problem space they can each control or monitor. Usually, each group member identifies a problem space and elements of that space, and additionally informs collaborators about the discrepancy between current and desired problem states (Larson and Christensen 1993).

Successful collaborative problem solving activities presuppose some kind of representation that is shared among participants. Research on so-called shared mental models has shown that teams demonstrate better problem solving performances if the individual problem representations (the individual mental models of the problem) are similar among group members (Klimoski and Mohammed 1994). Similarity among representations can be achieved through communication. In contrast to a shared mental model approach that just looks at similarities among individual representations, Roschelle and Teasley (1995) have proposed the concept of a joint problem space. This problem space is created and maintained through constant coordination and communication among collaborators, and serves as a basis for collaborative action.

Collaborators need a shared plan on how to achieve a goal state. Collaborative planning needs to include the management of resources. Research on transactive memory systems (Wegner 1986) has shown that groups benefit if members know who knows what or who has identified specific elements of the problem space in a group. In the case of groups composed of members with different problem-relevant knowledge (i.e., consistent with the requisite features of problems that might justify collaboration), the management of resources ideally takes into account that group members share all available information. The occurrence of information sharing is far from guaranteed: social psychological research has demonstrated that group members tend to mention shared information but neglect unshared information that is unique to only one group member (Stasser and Titus 1985). Resource allocation is not limited to knowledge. It also needs to include the identification of capacity to perform processing and the monitoring of processes.

Plans must be executed by the group. In some collaborative problem solving situations this requires an orchestrated effort by several group members in parallel. One of the pitfalls of collaborative action is that groups typically suffer from process losses (Steiner 1972), i.e., groups perform worse than they ideally could, given the members' abilities and resources. Process losses can be caused by group members' reduced task motivation (social loafing; Karau and Williams 1993), by additional social goals resulting from the group situation that are taking away resources from the task (Wittenbaum et al. 2004), and by reduced cognitive capacity due to the social situation (Diehl and Stroebe 1987).

Progress and courses of action must be evaluated, plans must be reformulated if necessary, and collaborators must decide on how to proceed. This again involves the risk of process losses. The analysis of monitoring activities can be informed by research on how groups implicitly and explicitly orchestrate decision making. For instance, groups can be characterised through their use of implicit social decision schemes like "truth wins", "majority wins", or "plurality wins" (Laughlin and Ellis 1986). Moreover, groups can be differentiated by their explicit timing of decision making procedures. While some groups start by making decisions and then seek evidence that supports their decisions, other groups demonstrate a deliberative approach that starts with the seeking of evidence and then converges on a decision (Hastie and Pennington 1991). More generally, the successful allocation of resources requires awareness of a group's progress concerning the problem it faces and the resources available within the group, and is facilitated by a shared understanding of the desired state (Peterson and Behfar 2005).

In this logical sequence of processes, participants externalise their individual problem solving processes, and coordinate these contributions into a coherent sequence of events. The degree to which this idealised sequence takes place in reality is unclear. In any given case, its occurrence will be dependent not only on the groups' dynamics but on the characteristics of the problem space.

Collaborative problem solving is not a uniform process but a complex, coordinated activity between two or more individuals. Consequently, efficient problem solving does not rely on a uniform skill but rather a set of distinguishable subskills which are deployed in accordance with situational needs. While the five processes mentioned above (problem identification, problem representation, planning, executing, monitoring) can serve to 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.

Collaborative Problem Solving Skills

Based on the literature in several research fields, the ATC21S™ project[1] has developed a framework consisting of a hierarchy of skills that play a pivotal role in collaborative problem solving. The identified skills must fulfill three criteria: (1) they must be measurable in large-scale assessment, (2) they must allow the derivation of behavioural indicators that (after some training) can be assessed by teachers in a classroom setting, and (3) they must be teachable. Only if these three conditions are met will collaborative problem solving skills become a part of learning diagnostics, both in everyday classroom practice and in large-scale assessment studies like PISA (OECD 1999).

The framework of collaborative problem solving skills proposed here is based on the distinction between two very broad skill classes: social skills and cognitive skills. Social skills constitute the "collaborative" part of "collaborative problem solving". They play an important role in collaborative problem solving but are also a feature of many other collaborative tasks. Cognitive skills constitute the "problem solving" part of "collaborative problem solving". These skills address typical cognitive issues of problem solving and have more in common with classical approaches to individual problem solving. To clarify this distinction it can be said that the social skills are about managing participants (including oneself), whereas cognitive skills are about managing the task at hand. In the following, both classes of skill are described and discussed in more detail.

[1] The acronym ATC21S™ has been globally trademarked. For purposes of simplicity the acronym is presented throughout the chapter as ATC21S.

Social Process Skills

In order to be successful in collaborative problem solving, individuals need a number of social skills to help them coordinate actions in synchrony with other participants. Our conceptualisation of social skills refers in particular to three classes of indicators that can be subsumed under the general rubric of social skills: participation, perspective taking, and social regulation (Table 2.1). Participation describes the minimum requirements for collaborative interaction. It refers to the willingness and readiness of individuals to externalise and share information and thoughts, and to be involved in the stages of problem solving (Stasser and Vaughan 1996). The concept of perspective taking skills refers to the ability to see a problem through the eyes of a collaborator (Higgins 1981). This can be extremely helpful, as it allows for smoother coordination among collaborators. Moreover, for particular types of tasks, perspective taking skills are essential, as a group cannot come to a solution unless its members have the capacity to understand the concrete situation their collaborators are in (e.g., Trötschel et al. 2011). Finally, the concept of social regulation skills refers to the more strategic aspects of collaborative problem solving (Peterson and Behfar 2005). Ideally, collaborators use their awareness of the strengths and weaknesses of all group members, to coordinate and resolve potential differences in viewpoints, interests and strategies.

Participation Skills

Many accounts in the learning sciences stress the importance of participation, albeit with slightly different focuses. According to socio-constructivist epistemologies, participation refers to the long-term process of becoming part of a community of practice (Lave and Wenger 1991). At first, learners take a peripheral role in a community (legitimate peripheral participation), but once they become more experienced as community members they take on more responsibilities. According to a cognitively and linguistically oriented epistemology, participation refers to the observable action of engaging in discourse. In this research tradition, Cohen (1994) suggested that the extent to which learners participate in a collaborative activity is the best predictor of individual learning outcomes, provided that a task is collaborative (i.e. it cannot be accomplished by division of labour alone) and provided that the problem is relatively ill-structured. Whichever epistemology is preferred, participation is regarded as a crucial concept in the learning sciences that constitutes or at least leads to learning.

Within the range of participation skills, our framework further distinguishes between three aspects: action, interaction, and task completion. “Action” refers to the general level of participation of an individual, irrespective of whether this action is in any way coordinated with the efforts of other group members. While most classical psychologists would argue that actions are just behavioural consequences of internal, cognitive processes, many learning scientists regard actions as the fundamental “carriers” of cognition (Hutchins 1995; Nardi 1996). Problem solvers

Table 2.1 Social skills in collaborative problem solving

Element	Indicator	Low	Middle	High
Participation				
Action	Activity within environment	No or very little activity	Activity in familiar contexts	Activity in familiar and unfamiliar contexts
Interaction	Interacting with, prompting and responding to the contributions of others	Acknowledges communication directly or indirectly	Responds to cues in communication	Initiates and promotes interaction or activity
Task completion/ perseverance	Undertaking and completing a task or part of a task individually	Maintains presence only	Identifies and attempts the task	Perseveres in task as indicated by repeated attempts or multiple strategies
Perspective taking				
Adaptive responsiveness	Ignoring, accepting or adapting contributions of others	Contributions or prompts from others are taken into account	Contributions or prompts of others are adapted and incorporated	Contributions or prompts of others are used to suggest possible solution paths
Audience awareness (Mutual modelling)	Awareness of how to adapt behaviour to increase suitability for others	Contributions are not tailored to participants	Contributions are modified for recipient understanding in the light of deliberate feedback	Contributions are tailored to recipients based on interpretation of recipients' understanding
Social regulation				
Negotiation	Achieving a resolution or reaching compromise	Comments on differences	Attempts to reach a common understanding	Achieves resolution of differences
Self evaluation (Metamemory)	Recognising own strengths and weaknesses	Notes own performance	Comments on own performance in terms of appropriateness or adequacy	Infers a level of capability based on own performance
Transactive memory	Recognising strengths and weaknesses of others	Notes performance of others	Comments on performance of others in terms of appropriateness or adequacy	Comments on expertise available based on performance history
Responsibility initiative	Assuming responsibility for ensuring parts of task are completed by the group	Undertakes activities largely independently of others	Completes activities and reports to others	Assumes group responsibility as indicated by use of first person plural

differ in the level of sophistication with which they act in a group. While some problem solvers do not become active at all, others become active once the environment is highly scaffolded (e.g. through explicit task instructions). Finally, the most sophisticated way of acting in a group is demonstrated by those who have the ability to perform actions even in the absence of instructional scaffolds.

"Interaction" refers to behaviour that demonstrates interaction with and responses to others. For instance, some learners are highly active in collaborative problem solving, but fail to respond to or coordinate with their collaborators. A higher level of interaction skill is exemplified by problem solvers who respond to cued interaction, e.g. by answering an inquiry from a collaborator. The highest level of interaction skill manifests itself if learners actively initiate coordination efforts, or prompt their collaborators to respond. Interaction among problem solvers is a minimum requirement for successful coordination (Crowston et al. 2006) and it is achieved through verbal and nonverbal means (Clark 1996).

"Task completion" skills refer to motivational aspects of participation and consequent perseverance on a task. Collaborative problem solvers differ in the degree to which they feel committed to the activity. Accordingly, they may enter the problem solving space but not be sufficiently engaged to remain actively involved, or at the other end of the spectrum, may persist in engagement as indicated by multiple attempts at tasks or by trying different strategies.

Perspective Taking Skills

While the quantity of participation is an important predictor of collaborative problem solving performance, perspective taking skills revolve more around the quality of interaction. Theoretically, perspective taking can be linked to constructs that stem from sub-disciplines as diverse as psychology of emotion, social psychology, and psycholinguistics, and consequently perspective taking encompasses affective, social-developmental, and linguistic aspects. Perspective taking is a multidimensional construct. On an affective level, perspective taking can be linked to the notion of empathy and the emotional understanding of, and identification with, others. More important in the current context, on a cognitive level, perspective taking is related to "theory of mind" concepts, and it describes the ability to understand a state of affairs from a different spatial or psychological perspective. If this ability is not in place, people are subject to egocentric bias, i.e. they expect others to be highly similar to themselves (Zuckerman et al. 1983). Perspective taking is often considered a core communicative competence (Weinstein 1969). Finally, a linguistic aspect of perspective taking refers to the ability to contextualise utterances of peers by reference to background information, but also the ability to tailor one's own utterances to the needs and intellectual capabilities of peer learners. This ability is often subsumed under the label of 'audience design' (Clark and Murphy 1982). It should be noted that while there is a general consensus among scholars that audience design is helpful to coordinate mutual activities, empirical evidence indicates that participants sometimes lack the ability or willingness to adapt to their communication partners (e.g. Horton and Keysar 1996).

The framework of collaborative problem solving skills distinguishes between two aspects of perspective taking skills: responding skills and audience awareness skills. Responding skills become apparent when problem solvers manage to integrate contributions of collaborators into their own thoughts and actions. For instance, problem solvers who rethink a problem representation based on evidence that was reported by a collaborator exhibit a high degree of responding skill. In contrast, ignoring contributions from others exemplifies a low degree of responding skill.

Audience awareness skills are constituted by the ability to tailor one's contributions to others (Dehler et al. 2011). Depending on variables like the amount of egocentric bias, problem solvers are more or less skilled in adapting their utterances to the viewpoints of others, or to making their actions visible and comprehensible to their collaborators. For example, imagine two problem solvers who are placed on different sides of a transparent screen. For a particular object on the left side from a problem solver's point of view, low audience awareness would be exhibited by referring to the object as being "on the left side". In contrast, higher audience awareness would be exemplified by referring to the object as being "on the right side" or even "on your right side".

To clarify the distinction between responding skills and audience awareness skills it can be said that the former involve the ability to be adaptive in one's internalisations of information (similar to Piaget's accommodation; Piaget and Inhelder 1962), whereas the latter involve the ability to be adaptive in one's externalisations of knowledge. The two aspects of perspective taking explicated in the current framework can thus be characterised respectively as *receptive* and *expressive*.

Social Regulation Skills

One of the main benefits of collaborating in a group is the potential diversity group members bring to their interactions. Different members have different knowledge, different expertise, different opinions, and different strategies. Evidence for the power of diversity has been found in the research of various disciplines that analyse group performance. For instance, in organisational psychology the concept of informational diversity among team members was identified as a key ingredient of team performance (De Wit and Greer 2008). The effects of diversity are particularly positive when group tasks require creativity and elaboration (van Knippenberg and Schippers 2007). In education, diversity among group members is considered to stimulate useful cognitive conflict (Doise and Mugny 1984), conceptual change (Roschelle 1992), or multiperspectivity (Salomon 1993). However, diversity per se is not in itself valuable and only becomes useful in collaboration when participants know how to deal with the diversity of viewpoints, concepts, and strategies under discussion (van Knippenberg et al. 2004). In other words, collaborative problem solvers need strategic skills to harness the diversity of group members, and they must employ mechanisms of social regulation and negotiation (Thompson et al. 2010) that act appropriately on group diversity. Groups have a tendency not to make use of the full potential of diversity (Hinsz et al. 1997). Among other things, dissenting information is often disregarded by individuals (confirmation bias; Jonas et al.

2001), shared information is preferred over unshared information (Stasser and Titus 1985), and minority viewpoints have less influence than majority viewpoints (Wood et al. 1994). If group members possess the skills to overcome biased information handling in groups and can regulate conflicts, they can fully exploit the benefits of diversity that their collaborators bring into the joint problem solving effort.

The framework of collaborative problem solving skills distinguishes four aspects that can be related to social regulation: metamemory, transactive memory, negotiation and initiative. The first two of these aspects refer to the ability to recognise group diversity, which breaks down into knowledge about oneself (metamemory; Flavell 1976), and knowledge about the knowledge, strengths, and weaknesses of one's collaborators (transactive memory; Wegner 1986). If these two skills are employed, collaborative problem solving groups will lay the groundwork to harness the power of group diversity.

The presence or absence of negotiation skills becomes apparent when conflicts arise among group members. These may be conflicts about how to represent a problem, about potential solution steps, about how to interpret evidence that is available to the group, or about the group's goals. In any of these cases, problem solvers must negotiate the steps and measures that accommodate the differences between individual approaches, for example by formulating compromises or by determining rank orders among alternative solution steps.

Finally, the term initiative skills refers to the responsibility that a problem solver experiences for the progress of the group. If this collective responsibility (Scardamalia 2002) is too low, lurking behaviour or disengagement from the task becomes likely, and it could be that the collaborative task becomes unsolvable. In contrast, higher responsibility is likely to contribute to better problem solving performance. While some problem solvers shun confrontation or even interaction by focusing on their individual solution attempts, others will take responsibility for working on a shared problem representation, developing a strategic plan towards a solution, and regularly monitoring activities on the group's progress.

If these different skills of social regulation are apparent in a group, the coordination of collaborative problem solving activities becomes much easier, and the potential diversity among group members will be exploited in highly beneficial ways.

Cognitive Process Skills

The effectiveness and efficiency of collaborative problem solving relies not only on social skills but also on cognitive skills. Cognitive skills of collaborative problem solving are highly similar to those skills that are conducive to individual problem solving, and they refer to the ways in which problem solvers manage the task at hand and the reasoning skills employed. The framework of collaborative problem solving categorises cognitive skills across planning, executing and monitoring, flexibility, and learning. Planning skills consist in an individual's capability to develop strategies based on plausible steps towards a problem solution (Miller et al. 1960).

In the case of collaborative problem solving, plans need to address a shared problem representation and provide the basis for an orchestrated and well coordinated problem solution (Weldon and Weingart 1993). While planning refers to prospective actions like building hypotheses, executing and monitoring is of a more retrospective nature. Problem solvers must interpret evidence, and must reflect on the appropriateness of planned and executed solution steps (Peterson and Behfar 2005). Monitoring is considered here as an individual-level skill, because it is more effective when it is done individually and externalised afterwards than when learners reflect jointly about the group process (Gurtner et al. 2007). This serves as a basis for the continuing adjustment of plans, thereby setting in motion a cyclical problem solving behaviour. Flexibility skills are demonstrated in the creativity that problem solvers exhibit when facing a particularly challenging part of a problem solution (Star and Rittle-Johnson 2008), but also include the way problem solvers react to ambiguous situations. These are particularly important if the problems are ill-defined and require some sort of inductive thinking. Finally, learning skills are demonstrated in the ability to learn during group interaction or as a consequence of group interaction. They lead to knowledge building. These four cognitive skill classes are elaborated in Table 2.2.

Table 2.2 Cognitive skills in collaborative problem solving

Element	Indicator	Low 0	Middle 1	High 2
Task regulation				
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
Sets goals	Sets a clear goal for a task	Sets general goal such as task completion	Sets goals for subtasks	Sets goals that recognise relationships between subtasks
Resource management	Manages resources or people to complete a task	Uses/Identifies resources (or directs people) without consultation	Allocates people or resources to a task	Suggests that people or resources be used
Flexibility and ambiguity	Accepts ambiguous situations	Inaction in ambiguous situations	Notes ambiguity and suggests options	Explores options
Collects elements of information	Explores and understands elements of the task	Identifies the need for information related to immediate activity	Identifies the nature of the information needed for immediate activity	Identifies need for information related to current, alternative, and future activity

(continued)

Table 2.2 (continued)

Element	Indicator	Low 0	Middle 1	High 2
Systematicity	Implements possible solutions to a problem and monitors progress	Trial and error actions	Purposeful sequence of actions	Systematically exhausts possible solutions
Learning and knowledge building				
Relationships (Represents and formulates)	Identifies connections and patterns between and among elements of knowledge	Focused on isolated pieces of information	Links elements of information	Formulates patterns among multiple pieces of information
Rules: "If … then"	Uses understanding of cause and effect to develop a plan	Activity is undertaken with little or no understanding of consequence of action	Identifies short sequences of cause and effect	Uses understanding of cause and effect to plan or execute a sequence of actions
				Plans a strategy based on a generalised understanding of cause and effect
Hypothesis "what if…" (Reflects and monitors)	Adapts reasoning or course of action as information or circumstances change	Maintains a single line of approach	Tries additional options in light of new information or lack of progress	Reconstructs and reorganises understanding of the problem in search of new solutions

Task Regulation Skills

"Planning" is one of the core activities of problem solving (Gunzelmann and Anderson 2003). On the basis of a (joint) problem space, planning involves the formulation of hypotheses concerning how to reach the goal, and the selection of steps that move the problem-solving process forward. Planning is a crucial meta-cognitive activity, as it requires problem solvers to reflect on their own (and others') cognitive processes (Hayes-Roth and Hayes-Roth 1979). We distinguish between four aspects of planning: problem analysis, goal setting, resource management and complexity. Planning begins with a *problem analysis*, an inspection of the individual or joint representation of a problem through which the task is segmented into sub-tasks with consequent sub-goals. Sub-tasks and sub-goals can not only make the problem solving process more tractable, they can also serve as important yardsticks to evaluate one's progress (i.e., monitoring). A good problem solver is able to *formulate specific goals* ("Next, we must move this block one tile to the left"), whereas lower sophistication is exhibited by formulating no goals or very vague ones ("We must try our best to change those blocks"). Research on teamwork has shown that goal specificity improves a group's performance (Weldon and Weingart 1993). The more a problem solver is inclined to set specific goals, the easier it is to assess and ultimately achieve them. Many collaborative problem solving tasks can only be accomplished if available resources are distributed properly. Resource management reflects the ability to plan how collaborators can bring their resources, their knowledge, or their expertise into the problem solving process. A low level of resource management skills is evident if a problem solver only plans with those resources that are available to herself. Suggesting that collaborators make use of specific resources indicates better resource management skills, whereas the highest skill level is exhibited when problem solvers explicitly decide on allocation of resources to people and/or task components. Therefore, an important aspect of planning is to *manage resources* that are available to oneself and to one's collaborators (Brown 1987). Finally, plans can differ in *complexity* or sophistication. This can best be described by reference to a chess match. If a piece is moved without prior reflection, planning complexity is low. If a sequence of moves is planned, and if potential counter moves are reflected in parallel plans of alternative routes, higher complexity in planning skill is demonstrated. To address these issues the framework of collaborative problem solving skills introduces the skill class of fluidity problems, which breaks down into two aspects: tolerance for ambiguity, and breadth. Different levels of ambiguity tolerance lead to different problem solving behaviours – some problem solvers become active only in unambiguous situations, some react to ambiguity by exploring the problem space, while problem solvers with high levels of ambiguity tolerance are likely to interpret ambiguous situations in a way that helps them in their decision making about the next solution step. As to breadth, a low skill level is displayed if problem solvers follow only a single approach of inquiry. A medium level of flexibility entails trying multiple approaches once an impasse is reached, or once new evidence is available via monitoring. And a high level of breadth leads to a re-organisation of problem representation or planning activities if progress through the problem space is impeded.

Problem solving is an activity that requires participants to cope with various barriers. For instance, most problems are inherently ambiguous because the best possible solution step is not always easily identifiable. Moreover, solution steps might lead to an impasse which represents a failure of the effort as it was originally planned. It is not uncommon for problem solvers to withdraw from a problem when they perceive roadblocks along the way to a solution. This can happen with all kinds of problems but it becomes particularly important for ill-defined problems that are ambiguous by definition. Tolerance for ambiguity (Norton 1975) is a characteristic of problem solvers that can help to overcome the barriers in problem solving activities. Moreover, good problem solvers are adept at changing plans in a flexible manner.

Research on human and machine problem solving has identified a number of recurring strategies that describe different approaches on how to tackle a problem.

For instance, one approach was termed ‘forward search’ (Newell and Simon 1972), and it can be characterised by taking a current problem state and identifying the most promising operator or move, thereby working towards the goal state. Variants of forward search include a breadth-first search (sequentially checking potential next moves) and depth-first search (following the most promising move until an impasse is reached). ‘Backward search’ through a problem space is the counterpart to forward search, and it starts with identifying the most likely or promising antecedent of a goal state, thereby working backwards through problem space. Backward search and forward search have been combined by Newell and Simon (1972), who have developed a means-ends-analysis based on the idea of selecting actions that minimise the difference between current state and goal state. This means-ends-analysis effectively comprises both forward search and backward search. However, while this and similar techniques can help to describe well-defined problems formally, they do not fully capture the complexity of ill-defined problems. For instance, many real-world problems are “wicked” because problem solvers lack necessary information (Van Gundy 1987). Realising that some crucial information is missing, and developing strategies on how to acquire this information, are important monitoring activities. In collaborative problem solving, this type of monitoring becomes essential, as different problem solvers typically have access to different types of information or have different means to access needed information (Larson and Christensen 1993).

Consequently, the framework of collaborative problem solving skills distinguishes between two “executing and monitoring” processes: information collection and systematicity. Information collection skill refers to the ability to identify what information is required and how and when it can be acquired. Some problem solvers lack the skills to identify the types of information required. Others will recognise the nature of the information needed, but only with regard to the current activity or problem state. Finally, a high level of these skills entails assessing the need for information with regard to current, alternative, and future problem states. Systematicity refers to the level of sophistication that a problem solver’s strategy exhibits. The most basic level of systematicity involves problem solving as a trial and error process. A medium level of systematicity is indicated by the use of forward search through a problem; whereas high systematicity can be identified when forward and backward search are combined through means-ends-analysis or similar techniques, followed by highly reflective monitoring activities.

Learning and Knowledge Building Skills

Brodbeck and Greitemeyer (2000) have characterised learning as a by-product of collaborative problem solving. Through progress in a collaborative problem solving task, individuals can learn about a content domain or about strategies and skills; they can also learn how to deal with impasses or how to coordinate, collaborate and negotiate with others. There are different ways to conceptualise learning, and the corresponding epistemologies for two of these have been described as participation and acquisition metaphors (Sfard 1998). The classical acquisition metaphor regards learning as the accumulation or restructuring of individual mental representations that leave measurable residues after a task is completed. In this case, the amount of learning can be measured through knowledge tests. In contrast, the participation metaphor is heavily influenced by situated cognition (Greeno 1998) and socio-culturalism (Vygotsky 1978), and regards learning as an activity rather than an outcome. The role of mental representations is downplayed and, according to this epistemology, knowledge is rather to be found in the environment (the task, the discourse, the artifact) than in the heads of learners. A particular view of learning that can be subsumed under the participation metaphor is knowledge building (Scardamalia 2002). According to this view, learning is a discursive process through which collaborators generate a network of ideas that build on each other. While the knowledge building epistemology seeks for learning during the process of collaborative problem solving, the acquisition metaphor of learning would assess learning through the transfer of skills or understandings.

The framework of collaborative problem solving skills touches on both these aspects, characterising the two as *knowledge building* and *learning*. Knowledge building is exemplified by the ability to take up ideas from collaborators to refine problem representations, plans, and monitoring activities. The highest level of knowledge building occurs in those problem solvers who are able to integrate and synthesise the input from collaborators (Scardamalia 2002) in the description and interpretation of a given problem. Learning is indicated by the ability to identify and represent relationships, understand cause and effect, and develop hypotheses based on generalisations. A low level of learning skills would be evident if the only knowledge that is extracted from a problem solving activity stems from information that was directly provided through instruction.

Griffin (2014) proposed a hierarchy of steps in problem solving which lead to knowledge building. At an initial level (beyond random guessing), students rely on identifying isolated elements of information. In a collaborative setting where information is unevenly and asynchronously distributed, these elements need to be shared. Problem solvers generally describe relationships or connections between elements of information (data) and make observations that form patterns, lending meaning to the problem space. At the next level of problem analysis, systematic observations of *cause and effect* enable players to formulate and discuss the potential of rules, either for the regulation of the task or for the manner of collaboration. At a more sophisticated level, rules are used to complete steps or parts of the problem solution. For the most difficult sub-tasks, more able students demonstrate an ability to generalise to a range of situations by setting and testing *hypotheses*, using a "What if…?" approach. An ordered progression, moving through pattern, rule and generalisation to hypothesis, can be developed by the collaborating partners and alternative solution options can be proposed and tested.

It is clear that there are overlapping cycles of cognitive processes across the general skill areas of task regulation – which includes planning, executing and monitoring, and comprehending complexity – and of knowledge building and learning. The essential difference between the two general areas consists in the use made by task regulation processes of the scoping of the problem space and the collection of information, which contrasts with the use of this information for extrapolation purposes in knowledge building and learning. For all the elements of the collaborative problem solving framework, the notions of teachability and learnability have been central to their conceptualisation. The rubrics in Table 2.2 give expression to the central place of this notion/these notions, and provide nutshell glimpses of the implications of the theoretical underpinnings of the construct for implementation in an assessment framework.

The debt of the presented framework to the work of Polya (1973), Mayer (1983), and the OECD PISA problem solving framework is substantial. The potential tension between a process approach to problem solving and a cognitive ability approach is evident in the long history concerning teachability of higher-order thinking processes. The ATC21S position, taking into account its assessment and teaching endeavour, is that the function of assessment is primarily to provide data to inform teaching. Consequently a process approach to collaborative problem solving is consistent with the project's primary goals. The extent to which individuals can be taught how to solve problems collaboratively is still unknown. It is clear that the distinct classes of sub-skills outlined in the framework can be taught. What is not so clear is whether an individual can be taught to draw on those sub-skills appropriately. It is at this point that the distinction between the process approach and a cognitive approach becomes the point of tension, and the focus for future research.

Assessment of Collaborative Problem Solving Skills

In order to assess problem solving skills in educational contexts, we must think about tasks that address the various skill classes described above. One of the decisions involved in identifying tasks relates to a trade-off between task realism and measurability. As to realism, collaborative problem solving can be found in many everyday activities: sitting together with a colleague and trying to format a software object; jointly developing a policy for student cafeteria use that takes into account the interests of various stakeholders; identifying a movie that is in line with the taste of a group of friends – all these are examples in which a group must identify a non-obvious solution that requires shared understanding and negotiation among collaborators. What these tasks often have in common is that they are ill-defined. For instance, the desired goal state cannot be clearly described (e.g. agreeing on a good cafeteria policy; finding a suitable movie). Furthermore, problems can be ill-defined because individuals and groups are not fully aware of the repertoire of actions that can lead them from the current state towards a goal state.

While many problems in real life are collaborative and ill-defined, the vast majority of research on problem solving has dealt with well-defined problems that are presented to individuals. A typical example for a well-researched problem is the "Tower of Hanoi" where individuals move disks according to specified rules in order to transform an original state into a well-defined goal state. Beginning with the seminal work by Newell and Simon (1972), an accumulation of research evidence has begun to show how individual problem solving behaviour can be understood and computationally modelled as the application of simple rules and heuristics. An advantage of these well-defined tasks is that their representational and computational dynamics are quite well understood. Consequently, there are agreed-upon standards for how to measure problem solving effectiveness.

The differences between real-world problems and problems as they are often analysed in psychological research raise the question of whether collaborative problem solving is best addressed by the use of well-defined or ill-defined tasks. Well-defined tasks allow for easier comparisons between different tasks and between different problem solvers, thereby providing the basis for the establishment of problem solving standards. Using well-defined tasks should also increase the teachability of collaborative problem solving, as the problem solving steps for well-defined tasks can be easily demonstrated, understood, adopted in the pursuit of alternative solution paths, or reflected upon. Therefore ATC21S has taken the approach that it is desirable for the design of collaborative problem solving tasks to begin with tasks that in some instances are designed for individual problem solving and transform these into collaborative tasks. For example, a typical approach to create collaborative (rather than cooperative) contexts is to introduce resource interdependence (Johnson et al. 1998). Modification of tasks can be implemented in this way to ensure that a task cannot be solved by any one individual working alone. The disadvantage of this approach is that it may not teach students to deal with truly ill-defined problems, since the constraints of the tasks are such that all resources are available, notwithstanding their lack of visibility.

Summary

With its wide applicability to real-life situations, collaborative problem solving – the joint and shared activity of transforming a current problem state into a desired goal state – can be regarded as one of the key skills in the 21st century. This chapter has proposed a framework that breaks down collaborative problem solving skills into a number of components. Most importantly, the social skills of collaboration can be distinguished from the cognitive skills of problem solving. Within these subgroups, certain skill aspects can be identified. The framework draws on research from several fields, and lays the ground for a deeper analysis of collaborative problem solving. One of the main purposes of this framework is to inform the design of collaborative problem solving tasks that touch on as many of the identified skill sets as possible. Once results from such tasks are available, testing of the theoretical hypotheses underlying the framework can take place in order to validate or refine the framework, thereby deepening our understanding of collaborative problem solving.

References

Brodbeck, F. C., & Greitemeyer, T. (2000). Effects of individual versus mixed individual and group experience in rule induction on group member learning and group performance. *Journal of Experimental Social Psychology, 36*(6), 621–648.

Brown, A. (1987). Metacognition, executive control, self-regulation, and other more mysterious mechanisms. In F. Reiner & R. Kluwe (Eds.), *Metacognition, motivation, and understanding* (pp. 65–116). Hillsdale: Erlbaum.

Clark, H. H. (1996). *Using language*. Cambridge, MA: Cambridge University Press.

Clark, H. H., & Murphy, G. L. (1982). Audience design in meaning and reference. *Advances in Psychology, 9*, 287–299.

Cohen, E. G. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research, 64*, 1–35.

Crowston, K., Rubleske, J., & Howison, J. (2006). Coordination theory: A ten-year retrospective. In P. Zhang & D. Galletta (Eds.), *Human-computer interaction in management information systems* (pp. 120–138). Armonk: M.E. Sharpe.

De Wit, F. R. C., & Greer, L. L. (2008). The black-box deciphered: A meta-analysis of team diversity, conflict, and team performance. In *Academy of Management best paper proceedings,* Anaheim.

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.

Diehl, M., & Stroebe, W. (1987). Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of Personality and Social Psychology, 53*(3), 497–509.

Dillenbourg, P., Baker, M., Blaye, A., & O'Malley, C. (1996). The evolution of research on collaborative learning. In E. Spada & P. Reiman (Eds.), *Learning in humans and machines: Towards an interdisciplinary learning science* (pp. 189–211). Oxford: Elsevier.

Doise, W., & Mugny, G. (1984). *The social development of the intellect*. Oxford: Pergamon Press.

Flavell, J. H. (1976). Metacognitive aspects of problem solving. In L. B. Resnick (Ed.), *The nature of intelligence* (pp. 231–236). Hillsdale: Erlbaum.

Greeno, J. G. (1998). The situativity of knowing, learning, and research. *American Psychologist, 53*(1), 5–26.

Griffin, P. (2014). Performance assessment of higher order thinking. *Journal of Applied Measurement, 15*(1), 1–16.

Gunzelmann, G., & Anderson, J. R. (2003). Problem solving: Increased planning with practice. *Cognitive Systems Research, 4*, 57–76.

Gurtner, A., Tschan, F., Semmer, N. K., & Nägele, C. (2007). Getting groups to develop good strategies: Effects of reflexivity interventions on team process, team performance, and shared mental models. *Organisational Behavior and Human Decision Processes, 102*(2), 127–142.

Hastie, R., & Pennington, N. (1991). Cognitive and social processes in decision making. In L. B. Resnick, J. M. Levine, & D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 308–327). Washington, DC: American Psychological Association.

Hayes-Roth, B., & Hayes-Roth, F. (1979). A cognitive model of planning. *Cognitive Science, 3*, 275–310.

Higgins, E. T. (1981). Role taking and social judgment: Alternative developmental perspectives and processes. In J. H. Flavell & L. Ross (Eds.), *Social cognitive development: Frontiers and possible futures* (pp. 119–153). Cambridge, UK: Cambridge University Press.

Hinsz, V. B., Tindale, R. S., & Vollrath, D. A. (1997). The emerging conception of groups as information processors. *Psychological Bulletin, 121*, 43–64.

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition, 59*, 91–117.

Hutchins, E. (1995). *Cognition in the wild*. Cambridge, MA: MIT Press.

Johnson, D., Johnson, R., & Holubec, E. (1998). *Cooperation in the classroom*. Boston: Allyn and Bacon.

Jonas, E., Schulz-Hardt, S., Frey, D., & Thelen, N. (2001). Confirmation bias in sequential information search after preliminary decisions: An expansion of dissonance theoretical research on selective exposure to information. *Journal of Personality and Social Psychology, 80*(4), 557–571.

Karau, S. J., & Williams, K. D. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology, 65*(4), 681–706.

Klimoski, R., & Mohammed, S. (1994). Team mental model: Construct or metaphor? *Journal of Management, 20*, 403–437.

Larson, J. R., Jr., & Christensen, C. (1993). Groups as problem-solving units: Toward a new meaning of social cognition. *British Journal of Social Psychology, 32*, 5–30.

Laughlin, P. R., & Ellis, A. L. (1986). Demonstrability and social combination processes on mathematical intellective tasks. *Journal of Experimental Social Psychology, 22*, 177–189.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. New York: Cambridge University Press.

Mayer, R. (1983). *Thinking, problem solving, cognition*. New York: W.H. Freeman and Company.

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behaviour*. New York: Holt, Rinehart & Winston.

Nardi, B. A. (1996). *Context and consciousness: Activity theory and human-computer interaction*. Cambridge, MA: MIT Press.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs: Prentice-Hall.

Norton, R. (1975). Measurement of ambiguity tolerance. *Journal of Personality Assessment, 39*(6), 607–619.

OECD. (1999). *Measuring student knowledge and skills: A new framework for assessment*. Paris: OECD.

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.

Piaget, J., & Inhelder, B. (1962). *The psychology of the child*. New York: Basic Books.

Polya, G. (1973). *How to solve it*. Princeton: Princeton University Press.

Roschelle, J. (1992). Learning by collaborating: Convergent conceptual change. *The Journal of the Learning Sciences, 2*(3), 235–276.

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

Salomon, G. (Ed.). (1993). *Distributed cognitions*. Cambridge: Cambridge University Press.

Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. In B. Smith (Ed.), *Liberal education in a knowledge society* (pp. 67–98). Chicago: Open Court.

Schoenfeld, A. H. (1999). Looking toward the 21st century: Challenges of educational theory and practice. *Educational Researcher, 28*, 4–14.

Schulz-Hardt, S., & Brodbeck, C. F. (2008). Group performance and leadership. In M. Hewstone, W. Stroebe, & K. Jonas (Eds.), *Introduction to social psychology: A European perspective* (4th ed., pp. 264–289). Oxford: Blackwell.

Sfard, A. (1998). On two metaphors for learning and the dangers of choosing just one. *Educational Researcher, 27*(2), 4–13.

Star, J. R., & Rittle-Johnson, B. (2008). Flexibility in problem solving: The case of equation solving. *Learning and Instruction, 18*(6), 565–579.

Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and Social Psychology, 48*(6), 1467–1478.

Stasser, G., & Vaughan, S. I. (1996). Models of participation during face-to-face unstructured discussion. In E. H. Witte & J. H. Davis (Eds.), *Understanding group behavior: Consensual action by small groups* (Vol. 1, pp. 165–192). Mahwah: Erlbaum.

Steiner, I. D. (1972). *Group processes and productivity*. New York: Academic.

Thompson, L. L., Wang, J., & Gunia, B. C. (2010). Negotiation. *Annual Review of Psychology, 61*, 491–515.

Trötschel, R., Hüffmeier, J., Loschelder, D. D., Schwartz, K., & Gollwitzer, P. M. (2011). Perspective taking as a means to overcome motivational barriers in negotiations: When putting oneself into the opponent's shoes helps to walk toward agreements. *Journal of Personality and Social Psychology, 101*(4), 771.

Van Gundy, A. B. (1987). *Creative problem solving: A guide for trainers and management*. Westport: Greenwood Press.

Van Knippenberg, D., & Schippers, M. C. (2007). Work group diversity. *Annual Review of Psychology, 58*, 515–541.

Van Knippenberg, D., De Dreu, C. K. W., & Homan, A. C. (2004). Work group diversity and group performance: An integrative model and research agenda. *Journal of Applied Psychology, 89*, 1008–1022.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Wegner, D. M. (1986). Transactive memory: A contemporary analysis of the group mind. In B. Mullen & G. R. Goethals (Eds.), *Theories of group behavior* (pp. 185–205). New York: Springer.

Weinstein, E. A. (1969). The development of interpersonal competence. In D. A. Goslin (Ed.), *Handbook of socialization theory and research* (pp. 753–775). Chicago: Rand McNally & Company.

Weldon, E., & Weingart, L. R. (1993). Group goals and group performance. *British Journal of Social Psychology, 32*(4), 307–334.

Wittenbaum, G. W., Hollingshead, A. B., & Betero, I. C. (2004). From cooperative to motivated information sharing in groups: Moving beyond the hidden profile paradigm. *Communication Monographs, 71*, 286–310.

Wood, W., Lundgren, S., Ouellette, J., Busceme, S., & Blackstone, T. (1994). Minority influence: A meta-analytic review of social influence processes. *Psychological Bulletin, 115*, 323–345.

Zuckerman, M., Kernis, M. H., Guarnera, S. M., Murphy, J. F., & Rappoport, L. (1983). The egocentric bias: Seeing oneself as cause and target of others' behavior. *Journal of Personality, 51*(4), 621–630.

Chapter 3
Assessment of Learning in Digital Networks

Mark Wilson and Kathleen Scalise

Abstract This chapter provides both conceptual and empirical information about the skillset of Learning in Digital Networks – Information Communications Technologies (LDN-ICT). Data are drawn from the pilot phase of the ATC21S™ project research and development process, and were collected from August to November 2011 across Australia, Finland, Singapore and the U.S.A. (The acronym ATC21S™ has been globally trademarked. For purposes of simplicity the acronym is presented throughout the chapter as ATC21S.) The paper concludes with a discussion of ideas about reporting and use of the consequent development progression which underlies the construct.

How to Assess Digital Learning

The ATC21S view of assessment is based on beliefs that the current practice of schooling is outmoded in the global working environment. For example, Cisco, Intel and Microsoft (2008) contrasted the typical context of student standardised assessment – having students take tests individually – with a situation in the outside world where people work both individually and in groups to share complimentary skills and accomplish shared goals. A second difference between schooling and the contemporary workplace arises from the nature of the test subjects themselves: today, school subjects are divided by disciplinary boundaries, but in the workplace this subject knowledge is applied across disciplinary boundaries in the process of solving real world problems. Moreover, these problems are not solvable by simply recalling facts or applying simple procedures, but are complex and ill-structured – and set in specific concrete contexts. Finally, the traditional "closed book" testing context is contrasted with a setting where people have access to a vast array of

M. Wilson (✉)
University of California, Berkeley, CA, USA
e-mail: MarkW@berkeley.edu

K. Scalise
University of Oregon, Eugene, OR, USA

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_3

information and technological tools, where the challenge is to strategically craft a solution (CIM 2008).

The ATC21S project commissioned a series of "white papers" to help establish this effort (now published in Griffin et al. 2012). Among them, the most important for this chapter are the "skills paper" (Binkley et al. 2012), and the "methodology paper" (Wilson et al. 2012). The first of these white papers lays out a scheme for encompassing and understanding the nature of these "new" skills and the ways in which they relate to traditional schools subjects. The scheme is referred to as "KSAVE," standing for Knowledge, Skills and Attitudes, Values and Ethics. Using this scheme as a basis, two particular twenty-first century skills were chosen for inclusion in an ATC21S assessment demonstration – collaborative problem solving and LDN-ICT. The latter is the focus of this chapter, and our particular slant on that will be described below. The second white paper lays out an approach to developing the new assessments, based on the insights of a groundbreaking U.S. National Research Council report (NRC 2001). The approach chosen is called the BEAR Assessment System (BAS: Wilson 2005, 2009a; Wilson and Sloane 2000), and it will not be detailed here other than to note that it is based on the following four principles.

Principle 1: Assessment should be based on a developmental perspective of student learning; the building block is a construct map of a progress variable that visualizes how students develop and how we think about their possible changes in response to items.

Principle 2: There must be a match between what is taught and what is assessed; the building block is the items design, which describes the most important features of the format of the items—the central issue, though, is how the items design results in responses that are related back to the levels of the construct map.

Principle 3: Teachers must be the managers of the system, with the tools to use it efficiently and effectively; the building block is the outcome space, or the set of categories of student responses that make sense to teachers.

Principle 4: There is evidence of quality in terms of reliability and validity studies and evidence of fairness; the building block is a measurement model that provides for multidimensional item responses and links over time, both longitudinally within cohorts and across cohorts.

(Wilson 2009b)

How these principles become embedded in the process and the product of the assessment development will be exemplified in the account below.

Learning in Networks: The Construct Map

The term "LDN-ICT" encompasses a wide range of subtopics, including learning in networks, information literacy, digital competence and technological awareness, all of which contribute to *learning to learn* through the development of enabling skills. In the current global economy, learning through digital networks, and the use of digital media, is becoming increasingly important in private life, in learning and in professional life. We predict that this aspect of learning will become very important in the future. We see this as being true at the individual level and local or regional levels as well as at international levels.

For the ATC21S project, the focus of LDN-ICT was on *learning in digital networks*, which was seen as being made up of four strands:

- Functioning as a consumer in networks;
- Functioning as a producer in networks;
- Participating in the development of social capital through networks;
- Participating in intellectual capital (i.e., collective intelligence) in networks.

In our view, LDN-ICT involves thinking across platforms and hardware implementations, and also thinking outside the computer itself, to other devices and uses of technology.

The Four Strands

The four strands mentioned above are seen as interacting together in the activity of learning in networks. They are conceptualised as parallel developments that are interconnected and make up that part of LDN-ICT that is concerned with learning in networks.

First, functioning as a Consumer in Networks (CiN) involves obtaining, managing and utilizing information and knowledge from shared digital resources and experts in order to benefit private and professional lives. It involves questions such as:

- Will a user be able to ascertain how to perform tasks (e.g. by exploration of the interface) without explicit instruction?
- How long will it take an experienced user to find an answer to a question using their mobile device?
- What arrangement of information on a display yields a more effective visual search?
- How difficult will it be for a user to find information on a website?

Second, functioning as a Producer in Networks (PiN) involves creating, developing, organizing and re-organizing information/knowledge in order to contribute to shared digital resources.

Third, developing and sustaining Social Capital through Networks (SCN) involves using, developing, moderating, leading and brokering the connectivities within and between individuals and social groups in order to marshal collaborative action, build communities, maintain an awareness of opportunities and integrate diverse perspectives at community, societal and global levels.

Fourth, developing and sustaining Intellectual Capital through Networks (ICN) involves understanding how tools, media and social networks operate and using appropriate techniques through these resources to build collective intelligence and integrate new insights into personal understandings.

In Tables 3.1, 3.2, 3.3, and 3.4, levels of these four strands have been described as hypothesized construct maps showing an ordering of skills or competencies involved in each. At the lowest levels of each are the competencies that one would expect to see exhibited by a novice or beginner. At the top of each table are the competencies that one would expect to see exhibited by an experienced person – someone who would be considered very highly literate in LDN-ICT. These construct maps are hierarchical in the sense that a person who would normally exhibit competencies at a higher level would also be expected to be able exhibit the competencies at lower levels of the hierarchy. The maps are also probabilistic in the sense that they represent different probabilities that a given competence would be expected to be exhibited in a particular context rather than certainties that the competence would always be exhibited.

These levels may be "staggered" in the sense that they have not been positioned on the same fixed scale for each strand. We see them as strands of the same broad construct – LDN-ICT – but the lower levels of one strand may be equivalent to the middle or even higher levels of other strands. This concept is represented in Fig. 3.1. It should also be noted that these construct maps were developed to encompass the full range of competencies within each strand rather than the range that one might expect to be exhibited by school students at middle and secondary levels. The question of targeting assessments to match what students can do is an empirical question to be determined through consultations with teachers and cognitive laboratories with students, as well as the results of pilot and field studies.

Table 3.1 Functioning as a Consumer in Networks (CiN)

	Consumer in networks
	Discriminating consumer
CiN3	Effectively judges credibility of sources/people
	Integrates information in coherent knowledge framework
	Conducts searches suited to personal circumstances
	Filters, evaluates, manages, organises and reorganises information/people
	Has little or no concept of credibility
	Selects optimal tools for tasks/topics
	Conscious consumer
CiN2	Selects appropriate tools and strategies (strategic competence)
	Constructs targeted searches
	Compiles information systematically
	Knows that credibility is an issue (web pages, people, networks)
	Emerging consumer
CiN1	Performs basic tasks
	Has no concept of credibility
	Searches for pieces of information using common search engines (e.g. movie guides)
	Knows that tools exist for networking (e.g. Facebook)

Table 3.2 Functioning as a Producer in Networks (PiN)

	Producer in networks
	Creative producer
PiN3	Possesses team-situational awareness in process
	Optimises assembly of distributed contribution to products
	Extends advanced models (e.g. business models)
	Produces attractive digital products using multiple technologies/tools
	Chooses among technological options for producing digital products
	Functional producer
PiN2	Establishes and manages networks & communities
	Possesses awareness of planning for building attractive websites, blogs, games
	Organizes communication within social networks
	Develops models based on established knowledge
	Develops creative & expressive content artifacts
	Possesses awareness of security & safety issues (ethical and legal aspects)
	Uses networking tools and styles for communication among people
	Emerging producer
PiN1	Produces simple representations from templates
	Starts an identity
	Uses a computer interface
	Posts an artifact

Table 3.3 Developing Social Capital through Networks (SCN)

	Developer of social capital
	Visionary connector
SCN4	Takes a cohesive leadership role in building a social enterprise
	Reflects on experience in social capital development
	Proficient connector
SCN3	Initiates opportunities for developing social capital through networks (e.g. support for development)
	Encourages multiple perspectives and supports diversity in networks (social brokerage skills)
	Functional connector
SCN2	Encourages participation in and commitment to a social enterprise
	Possesses awareness of multiple perspectives in social networks
	Contributes to building social capital through a network
	Emerging connector
SCN1	Participates in a social enterprise
	Is an observer or passive member of a social enterprise
	Knows about social networks

Table 3.4 Developing Intellectual Capital through Networks (ICN)

	Participant in intellectual capital (collective intelligence)
	Visionary builder
ICN4	Questions existing architecture of social media and develops new architectures
	Functions at the interfaces of architectures to embrace dialogue
	Proficient builder
ICN3	Understands and uses architecture of social media such as tagging, polling, role-playing and modelling spaces to link to knowledge of experts in an area
	Identifies signal versus noise in information
	Interrogates data for meaning
	Makes optimal choice of tools to access collective intelligence
	Shares and reframes mental models (plasticity)
	Functional builder
ICN2	Acknowledges multiple perspectives
	Uses thoughtful organization of tags
	Understands mechanics of collecting and assembling data
	Knows when to draw on collective intelligence
	Shares representations
	Emerging builder
ICN1	Possesses knowledge of survey tools
	Is able to make tags
	Posts a question

Learning in Networks: Three Scenarios

The Berkeley Evaluation and Assessment Research (BEAR) Center at UC Berkeley developed three scenarios in which to place tasks and questions that could be used as items to indicate where a student might be placed along each of the four strands. Each scenario was designed to address more than one strand, but there were different emphases in how the strands were represented among the scenarios. Where possible, we took advantage of existing web-based tools for instructional development. These are each briefly described below.

Arctic Trek

One potential mechanism for the assessment of student ability in the learning network aspect of LDN-ICT is to model assessment practice through a set of exemplary classroom materials. The module that has been developed is based on the Go North/Polar Husky information website (www.polarhusky.com) run by the University of Minnesota (see Fig. 3.2). The Go North website is an online adventure learning project based around arctic environmental expeditions. The website is a learning hub with a broad range of information and many different mechanisms to support

CiN

Level	What the student knows
3	Discriminating consumer
2	Conscious consumer
1	Emerging consumer

PiN

Level	What the student knows
3	Creative producer
2	Functional producer
1	Emerging producer

SCN

Level	What the student knows
4	Visionary connector
3	Proficient connector
2	Functional connector
1	Emerging connector

ICN

Level	What the student knows
4	Visionary builder
3	Proficient builder
2	Functional builder
1	Emerging Builder

Fig. 3.1 The four strands of LDN-ICT, represented as a four-part learning progression

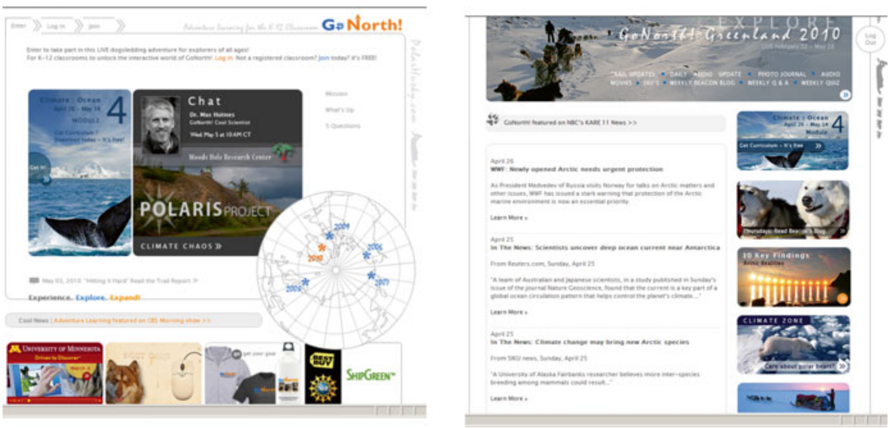


Fig. 3.2 Two screen-shots from the *Go-North!* Website

networking with students, teachers and experts. LDN-ICT resources developed for this module focus mainly on the functioning as a Consumer in Networks strand. The tour through the site for the ATC21S demonstration scenario is conceived as a "collaboration contest," or virtual treasure hunt. The Arctic Trek scenario views social networks through LDN-ICT as an aggregation of different tools, resources and people that together build community in areas of interest. In this task, students in small teams ponder tools and approaches to unravel clues through the Go North site by touring scientific and mathematics expeditions of actual scientists. The task helps teachers model ways to integrate technology across different subjects. It also shows how the Go North site focuses on space to represent itself, and how this can be combined with tools that utilize texting, chat and dialogue as forms of LDN-ICT.

Webspiration

In the second demonstration task, framed as part of a poetry work unit, students of ages 11–15 read and analyse well-known poems. In a typical school context, we might imagine that a teacher notices that his or her students are having difficulty articulating the moods and meanings of some poems – in traditional teacher-centered instruction on literature the student role tends to be passive. Often, teachers find that students are not spontaneous in their responses to poems but tend to wait to hear what the teacher has to say, and then agree with it. To help encourage students to formulate their own ideas on the poems, we use a collaborative graphic organiser through the Webspiration online tool. The teacher directs the students to use Webspiration to create an idea map – collaboratively using the graphic organizer tools – and to analyze each poem they read. Students submit their own ideas and/or build on classmate thoughts. Figure 3.3 shows a sample screen from the computer module.

Second Language Chat

This scenario was developed as a peer-based second language learning environment through which students interact in learning. Developing proficiency in a second language (as well as in the mother tongue) requires ample opportunities to read, write, listen and speak. This assessment scenario asks students to set up a technology/network-based chat room, invite participants and facilitate a chat – in two languages. It also involves evaluating the chat and working with virtual rating systems and online tools such as spreadsheets. The welcome screen for this scenario is shown in Fig. 3.4. "Conversation partner" language programs such as this have sprung up worldwide in recent years. They bring together students wishing to practise a language with native speakers, often in far-flung parts of the world. The cultural and linguistic exchanges that result demonstrate how schools can dissolve the

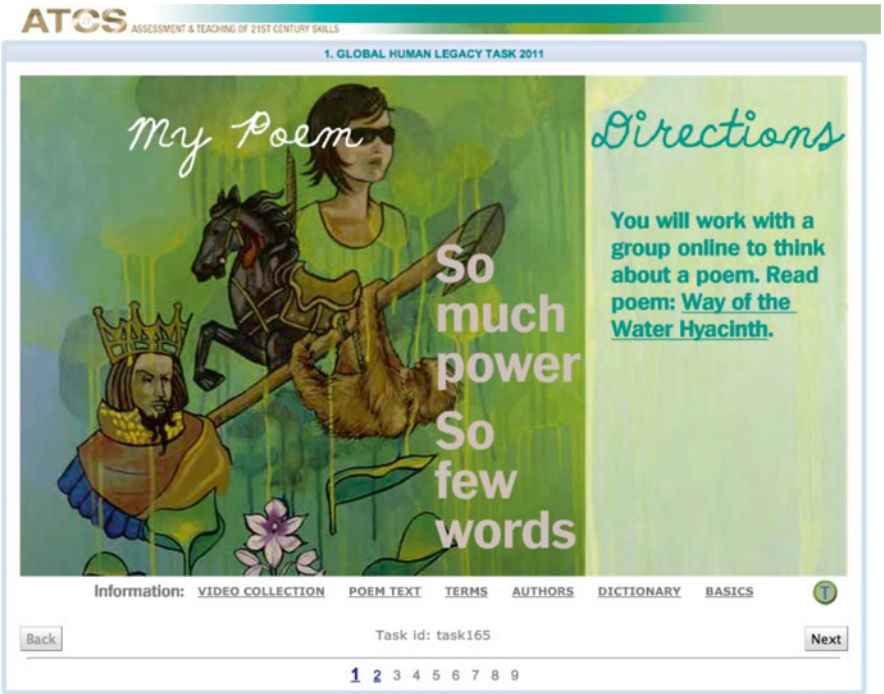


Fig. 3.3 A sample page from the Webspiration scenario

physical boundaries of walls and classrooms. They also tap rich new learning spaces through the communication networks of LDN-ICT. This task shows how they can also provide ample assessment opportunities in digital literacy.

Sample Tasks from Arctic Trek

The welcome screen from Arctic Trek is shown in Fig. 3.5. The student goal is to discover answers to 6 questions and each student must join a team to do this (see Fig. 3.6). Once the team is assembled, it must assign roles to each team member (Figs. 3.7 and 3.8). There is also a Team Notebook where its findings will be recorded (Fig. 3.9). The team then finds out about the contest (Fig. 3.10). There is a practice first – members must use the web resources listed in the right-hand panel to answer the question (Fig. 3.11). If a student cannot write down a response, then he or she can request a hint (and this can be repeated). The hints appear at the bottom of the screen (Fig. 3.12). If the hints are not enough (and eventually they do virtually tell the student what to do) then the student may request teacher assistance by hitting the “T” button at the bottom right-hand corner, but when that happens, the teacher must fill in an information box (Fig. 3.13). A real task is shown (partially)

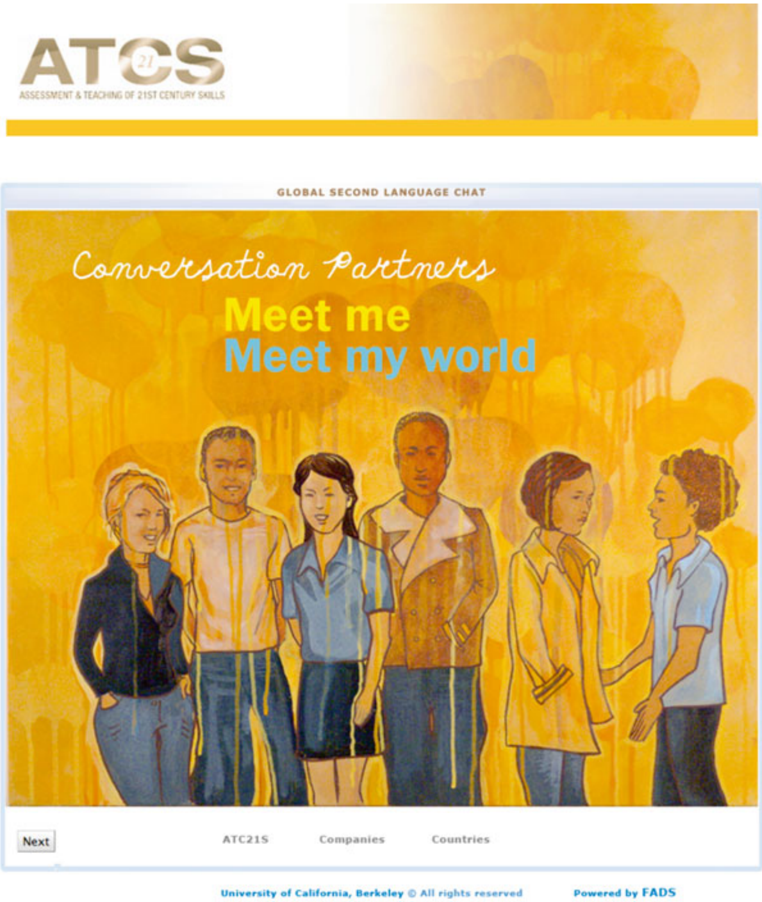


Fig. 3.4 The welcome page from the Two-language chat scenario

in Fig. 3.14 – student foraging in an online display. Here the student has been asked to examine a map that shows where polar bears are found and must describe the way the information is conveyed on the map.

Samples of student Team Notebooks are shown in Figs. 3.15 and 3.16. The first, Notebook A (from a group of 15-year-olds), shows clear role-selection, responses to the clues and explanations of response choice. The second, Notebook B (from a group of 11-year-olds), shows a very different team response – mainly arguing about roles. In this case, the responses to the questions are missing. Samples of data codes from two different teams are shown in Figs. 3.17 and 3.18. In the top panel of Fig. 3.17, the data codes show that Team #1 (a) successfully retrieved the team code, and (b) successfully accessed the shared notebook. They also show that (c) the team successfully assigned team roles, and there was consensus among the team members about those roles. In the lower panel of Fig. 3.17, the data codes show that

Fig. 3.5 The welcome screen from *Arctic Trek*

Fig. 3.6 Meeting the team

Team #1 (d) gave the correct answer for the number of colours, and (e) correctly listed the colours, and noted the issue about missing data. It also shows (f) that they used no hints or teacher assistance, and (g) that their self-evaluation of their collaboration was "Good." The account of Team #2, as shown in the data codes, is very different. In the top panel of Fig. 3.18, the data codes show that Team #2 (a) did not retrieve the team code, but (b) did successfully access the shared notebook.

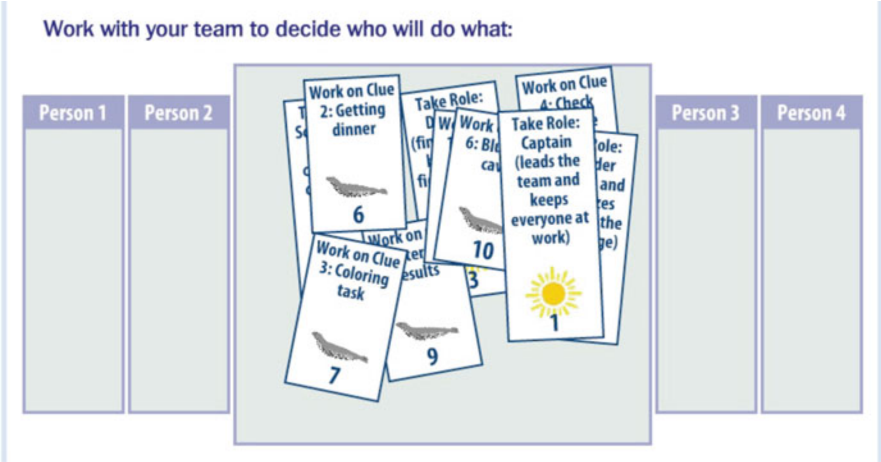


Fig. 3.7 Setting up the team roles

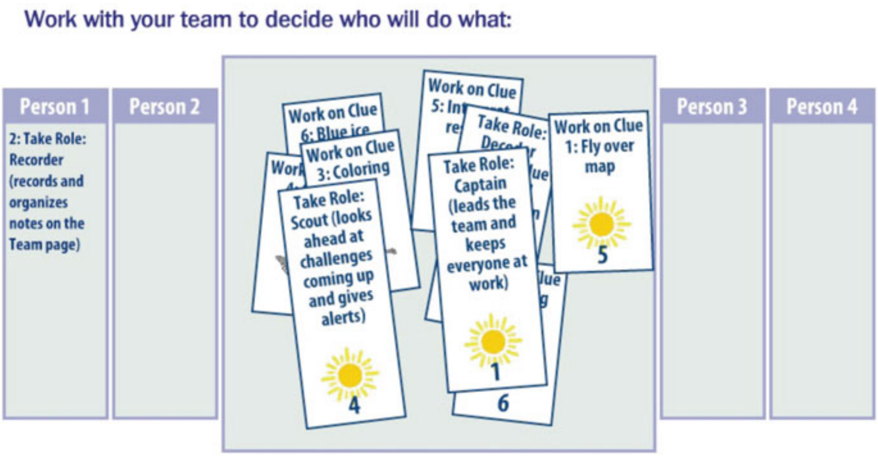


Fig. 3.8 Person 1 has been assigned as "Recorder"

They also show that (c) the team was unsuccessful in assigning team roles, and that there was no consensus among the team members about those roles. In the lower panel of Fig. 3.18, the data codes show that Team #2 (d) gave the correct answer for the number of colors, and (e) they compared answers, but did not note the issue about missing data. It also shows (f) that they used no hints or teacher assistance, and (g) that their self-evaluation of their collaboration was "Great" because "everyone in my group agreed."

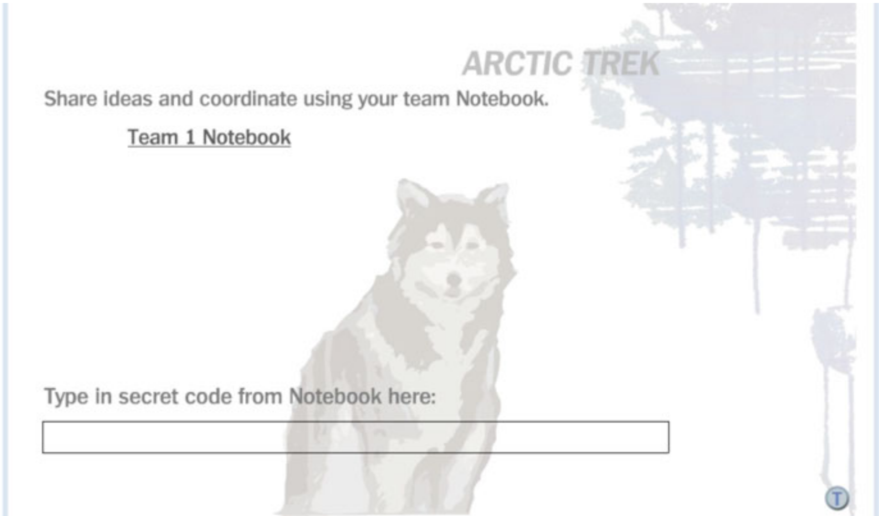


Fig. 3.9 Setting up the shared Team Notebook

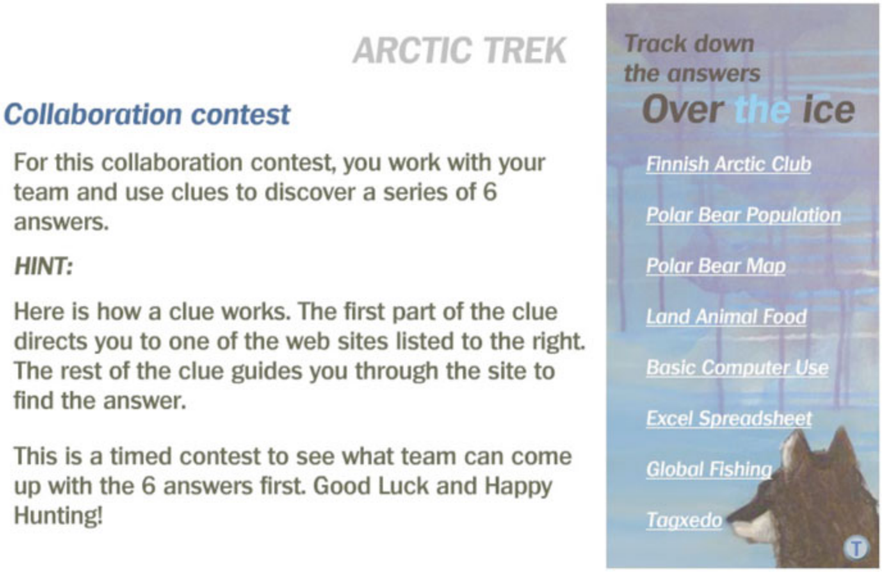


Fig. 3.10 The collaboration contest

The Outcome Space for the Three Scenarios

Each item was developed to target one or more of the four strands, and the expected range of levels that would be represented in the item responses was also noted. Where the responses are selected from a fixed set (as in a multiple-choice item), this

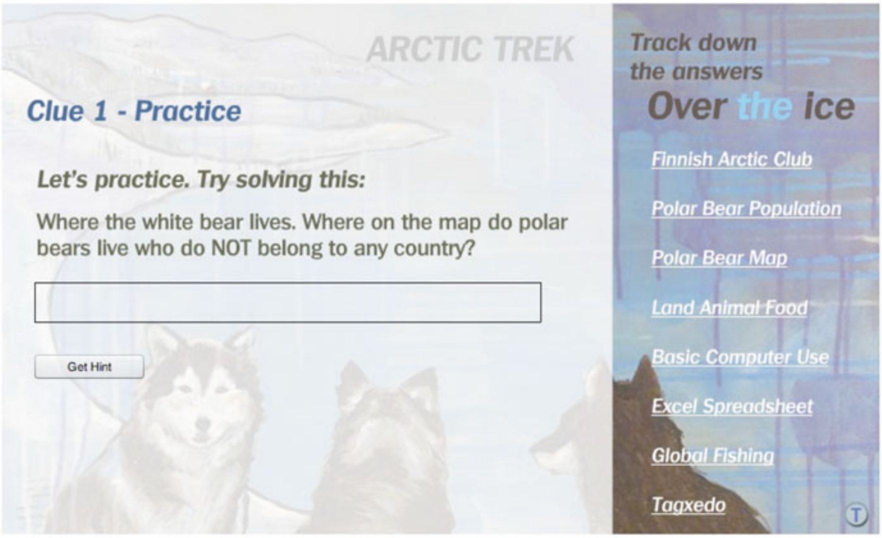


Fig. 3.11 An opportunity to practice

Fig. 3.12 A hint

can be planned ahead of time, but for open-ended items, this is something that needs to be empirically investigated. The tabulation is shown in Table 3.5. As can be seen, the first three levels were reasonably well covered, but Level 4, which we expect to see seldom for students in this population, had only one instance.

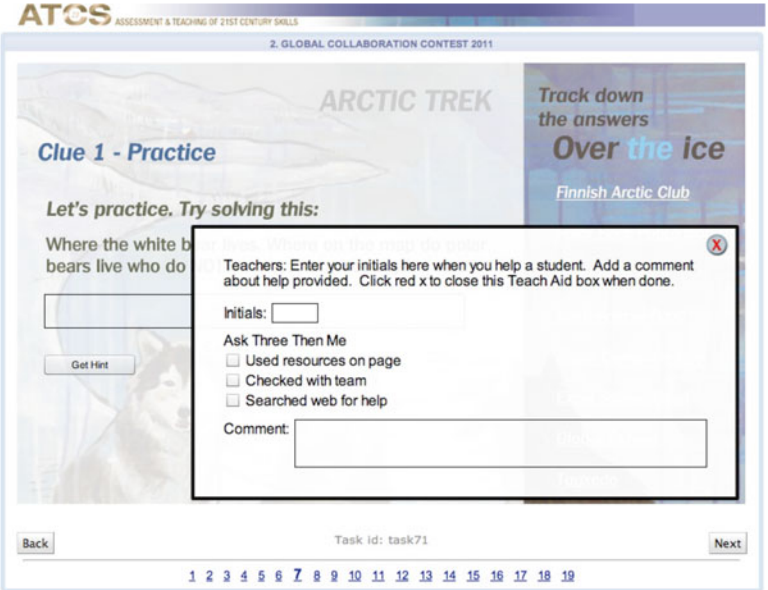


Fig. 3.13 The teacher aid box

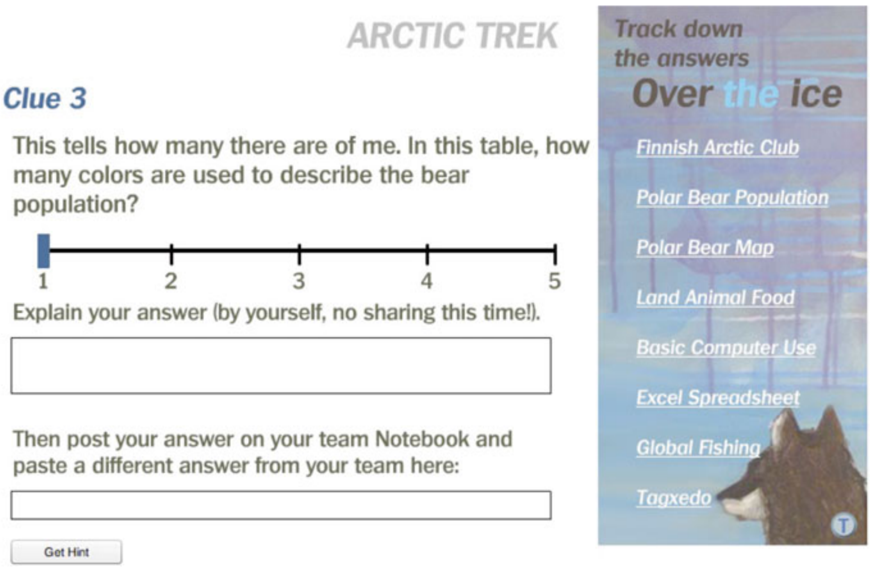


Fig. 3.14 The third clue

Roles
(1)David: Captain, Decoder
(2)Stephenie:Work on clue 1,2,3
(3)Xinyi: Work on clue 4,5,6
(4)Amanda: Scout, Recorder

Clues
Clue 1:
Where the white bear lives. Where on the map do polar bears live who do NOT belong to any country?
Polar bears live in North Pole

Clue 2:
Arctic Fox

Clue 3:
Answers of person:
(1)There are 5 colours of red, orange, light green, dark green and yellow with one extra colour white to represent data deficiency
(2) 5
(3) 5
(4) 5

Fig. 3.15 Sample notebook A

first q is challenging
i can i be the decoder ? there are two colors used in the table . page 15 is very hard !!!!!!!!!!!!!!!!!!!!!!!!!!
Your team's SECRET CODE: XF9
i don't know. i think may should be captin do you guys agree it's jas do you want to be a scout i will jas hmmm agggggggg! ok lets work this out lets vote!me me capinok i think may should be ca
year i know
how do you want to work it out whos going to be captin
yes
WHAT CAN I DO
Jonathan should be scout
who should should be the decoder
i will ally
i havent got a job
w
who is going to be the captain what does everybody want to be who is the captain may 2 because thereare two colours on the table
i will be capten
than i will be capten

Fig. 3.16 Sample notebook B

	Sc02 screen5v2 - Wave Intro DifferT		
a	ArcticB2	N/A	SJ3
	ArcticB2_PageLinks	N/A	Team 11 Notebook
	ArcticB2_taid_initials	0 / 0	
	ArcticB2_taid_comment	0 / 0	
	ArcticB2_taid_page	0 / 1	unchecked
	ArcticB2_taid_team	0 / 1	unchecked
	ArcticB2_taid_web	0 / 1	unchecked
	Sc02 assignRole		
	WebE1	1 / 1	Person 1
	WebE2	1 / 1	Person 4
	WebE3	0 / 1	Person 1
c	WebE4	0 / 1	Person 4
	WebE5	0 / 1	Person 2
	WebE6	0 / 1	Person 2
	WebE7	0 / 1	Person 2
	WebE8	0 / 1	Person 3
	WebE9	0 / 1	Person 3
	WebE10	0 / 1	Person 3
	WebE_Links	N/A	
	WebE_taid_initials	0 / 0	
	WebE_taid_comment	0 / 0	
	WebE_taid_page	0 / 1	unchecked
	WebE_taid_team	0 / 1	unchecked
	WebE_taid_web	0 / 1	unchecked

d	Sc02-Screen14		
	ArcticC5	0 / 1	5
	ArcticC6	N/A	There are main 5 colours: red, orange, light green, dark green and yellow with one extra colour white to represent data deficient
e	ArcticC7	N/A	It should be 5, because the colours used are meant to describe the population of polar bears. White, is used to represent insufficient data
	ArcticC7_Hints	0 / 0	0
	ArcticC_Links	0 / 1	
	ArcticC_taid_initials	0 / 0	
	ArcticC_taid_comment	0 / 0	
	ArcticC_taid_page	0 / 1	unchecked
	ArcticC_taid_team	0 / 1	unchecked
	ArcticC_taid_web	0 / 1	unchecked
g	Sc02-Screen15		
	ArcticC8	1 / 1	Good
	ArcticC9	0 / 1	5
	ArcticC10	N/A	Because the number line limits the number of colours to 5 and we have to ignore the extra colour white.
	ArcticC_Links	N/A	
	ArcticC_taid_initials	0 / 0	
	ArcticC_taid_comment	0 / 0	
	ArcticC_taid_page	0 / 1	unchecked
	ArcticC_taid_team	0 / 1	unchecked
	ArcticC_taid_web	0 / 1	unchecked

Fig. 3.17 Sample collaboration #1

Samples of teachers in Australia, Finland, Singapore and the United States were asked to provide feedback about draft tasks for LDN-ICT. Those teachers were provided with access through a teacher interface and for each set of tasks they were asked a set of questions to consider. These questions included:

For Webspiration

What skills or capabilities do you think the tasks are targeting?

Considering the capabilities of your students, are there any questions or activities that should be eliminated from this scenario, for students of specified ages (11, 13 and 15 years).

Sc02 screen5v2 - Wave Intro DifferT		
ArcticB2	N/A	
ArcticB2_PageLinks	N/A	Team 1 Notebook
ArcticB2_taid_initials	0 / 0	
ArcticB2_taid_comment	0 / 0	
ArcticB2_taid_page	0 / 1	unchecked
ArcticB2_taid_team	0 / 1	unchecked
ArcticB2_taid_web	0 / 1	unchecked
Sc02 assignRole		
WebE1	1 / 1	Person 1
WebE2	0 / 1	Person 2
WebE3	0 / 1	Not Sorted
WebE4	0 / 1	Person 4
WebE5	0 / 1	Person 3
WebE6	0 / 1	Not Sorted
WebE7	0 / 1	Not Sorted
WebE8	0 / 1	Not Sorted
WebE9	0 / 1	Not Sorted
WebE10	0 / 1	Not Sorted
WebE_Links	N/A	
WebE_taid_initials	0 / 0	
WebE_taid_comment	0 / 0	
WebE_taid_page	0 / 1	unchecked
WebE_taid_team	0 / 1	unchecked
WebE_taid_web	0 / 1	unchecked

Sc02-Screen14		
ArcticC5	0 / 1	5
ArcticC6	N/A	there are 5 because are five different colors even though there shades are alike
ArcticC7	N/A	everyone said 5
ArcticC7_Hints	0 / 0	0
ArcticC_Links	1 / 1	Land Animal Food,Basic Computer Use,Polar Bear Population
ArcticC_taid_initials	0 / 0	
ArcticC_taid_comment	0 / 0	
ArcticC_taid_page	0 / 1	unchecked
ArcticC_taid_team	0 / 1	unchecked
ArcticC_taid_web	0 / 1	unchecked
Sc02-Screen15		
ArcticC8	1 / 1	Great
ArcticC9	0 / 1	5
ArcticC10	N/A	because everyone in my group agreed
ArcticC_Links	N/A	
ArcticC_taid_initials	0 / 0	
ArcticC_taid_comment	0 / 0	
ArcticC_taid_page	0 / 1	unchecked
ArcticC_taid_team	0 / 1	unchecked
ArcticC_taid_web	0 / 1	unchecked
Sc02-Screen17		

Fig. 3.18 Sample collaboration #2 (Note that the locations of points "a" through "g" in the text are equivalent to those for Fig. 3.17)

For Arctic Trek

Identify and write down two clues to retain and two clues to eliminate from the task for students of specified ages.

Table. 3.5 The number of data points from each scenario and their planned allocation to the levels from each strand

Levels[a] (progressive)	ICT literacy – learning in digital networks				
	Construct/learning outcomes				
	Consumer	Producer	Social capital	Intellectual capital	Total
Level 4	N/A	N/A	Web 0	Web 0	Web 0
			Arctic 1	Arctic 0	Arctic 1
			2LChat 0	2LChat 0	2LChat 0
Level 3	Web 0	Web 0	Web 0	Web 10	Web 10
	Arctic 2	Arctic 2	Arctic 6	Arctic 2	Arctic 12
	2LChat 0	2LChat 0	2LChat 1	2LChat 1	2LChat 2
Level 2	Web 8	Web 4	Web 7	Web 6	Web 25
	Arctic 6	Arctic 16	Arctic 0	Arctic 7	Arctic 29
	2LChat 0	2LChat 8	2LChat 6	2LChat 0	2LChat 14
Level 1	Web 2	Web 4	Web 1	Web 2	Web 9
	Arctic 2	Arctic 0	Arctic 0	Arctic 2	Arctic 4
	2LChat 2	2LChat 6	2LChat 6	2LChat 0	2LChat 14
Total	Web 10	Web 8	Web 8	Web 18	Web 44
	Arctic 10	Arctic 18	Arctic 7	Arctic 11	Arctic 46
	2LChat 2	2LChat 14	2LChat 13	2LChat 1	2LChat 30

[a]Some CR items (constructed response) will measure up through the listed level (listed level is top score)

For Language Chat

At what age do you believe native speakers would be able to learn and use a rating system?
At what age would native speakers be able to facilitate a chat topic?
Suggest a chat topic for language learners at the selected age that has the potential to engage them.

Cognitive laboratories, which involve small samples of students who attempt the tasks and respond to questions about them, were also carried out in the four countries on all three task demonstrations. Information from these two sources contributed to the final editing of the tasks, and to the compilation of the information in Table 3.5.

Results from the Pilot Study

In the pilot study, two of the three scenarios were selected for further studies with students: the science/math Arctic Trek collaboration contest and the Webspiration shared literature analysis task. These were identified by participating countries as the most desirable to pilot at this time, for several reasons. These included that they were more aligned with traditional school systems in the countries, which rarely used cross-country chat tools in the classroom but sometimes did employ math simulations and online scientific documents as well as graphical and drawing tools for student use. By contrast, the third task – the Second Language Chat – was described by participating countries, teachers and schools as a forward-looking, intriguing scenario, but farther away on the adoption curve for school-based technology.

Not all of the planned automated scoring and data analysis for the items in the two piloted scenarios has been applied to this data set, as the total number of cases was too small for the empirically-based scoring to be successfully calibrated. This will be completed when larger data sets are available. Each of the two scenarios was presented in three forms, for 11, 13 and 15 year-olds respectively, with a subset of common items across the three forms. Due to the nature of the pilot study data design, results for the two scenarios are reported separately. The data were analysed using a partial credit item response model (Masters 1982), and the estimation software was ConQuest 2.0 (Wu et al. 2007).

For the Webspiration scenario, 176 cases were collected across Australia, Finland, Singapore and the U.S.A. Approximately 90 % of the items were auto-scored and 10 % were hand-scored (by trained scorers using a common scoring guide). There are 61 items in the three forms, and 16 are common across all forms. Approximately 10 % of the items showed significant misfit – these items will be retained for further examination in the field test. The reliability was estimated at 0.93 using the EAP formulation (Wu et al. 2007). The Wright Map, showing how items compare to students on the composite Learning in Networks latent variable is shown in Fig. 3.19.

Note that, due to the small number of cases available at this point, the four strands are all mapped onto the same composite variable. With a greater number of sample cases, this will be investigated using a multidimensional model. The map shows that students are reasonably well-matched by the range of item difficulties. Examination of the match between empirical locations of the item responses and the four strand construct maps resulted in a segmentation of the variable into five levels that correspond quite well with the planned levels.

The five levels are indicated by the alternating yellow and white bands in Fig. 3.19. The lowest two bands are associated with the first level of the strand construct maps. In the lowest band, students are required to move information (e.g., cut/paste, drag/drop, texting), ask simple questions, and begin to use rankings to arrange crowd-sourced information. In the second band, they correctly access team and individual pages and begin to discriminate among the crowd-sourced information provided. The third band is associated with the second levels of the strand construct maps: students search for targeted information, create links to displayed ideas, and use context to discriminate crowd-sourced information. The fourth band also is associated with the second level of the strand construct maps: students access digital tools and resources available in the environment, and select/share tagged ideas. The highest band is associated with the third level of the strand construct maps: students create explanations in new media and use tools to share products with others in new

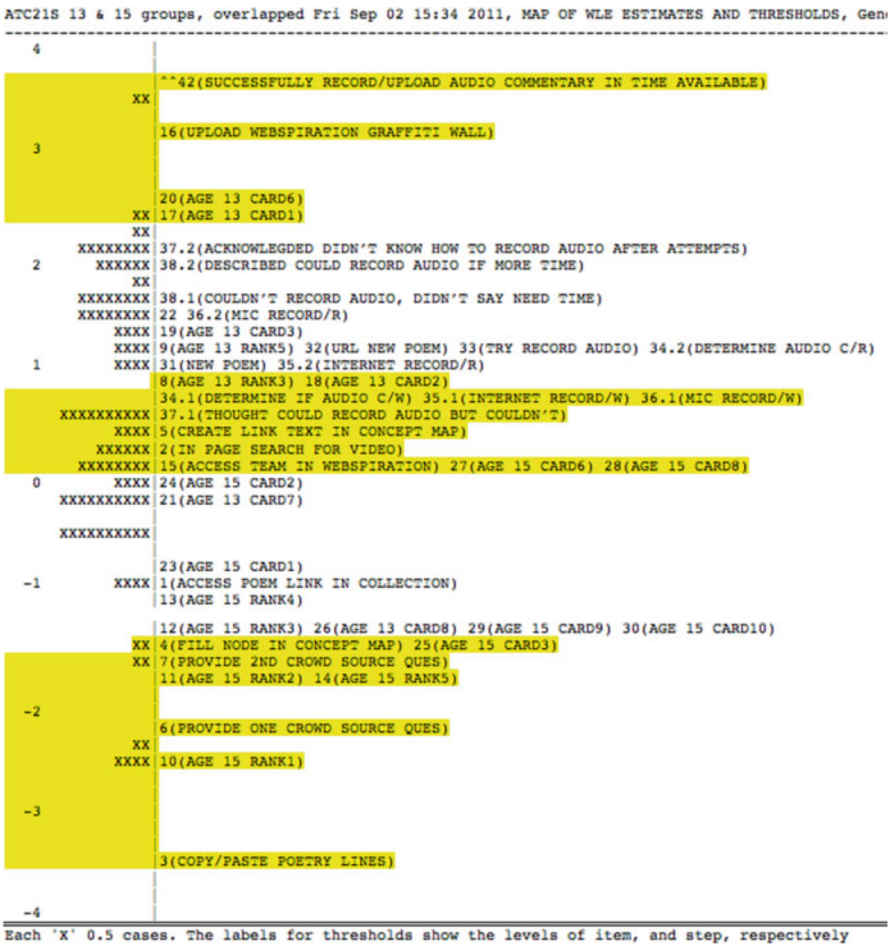


Fig. 3.19 Variable map for composite construct using the Webspiration scenario

interfaces. As expected, this highest level is rarely seen in the data for the sample population assessed in the tasks to date.

For the Arctic Trek scenario, 135 cases were collected across Australia, Finland and the U.S.A. Approximately 84 % of the items were auto-scored and 16 % were hand-scored (again, by trained scorers using a common scoring guide). There are 25 items in the three forms, and 20 are common across all forms. Approximately 8 % of the items showed significant misfit – these items will be retained for further examination in the field test. The reliability was estimated at 0.88 using the EAP formulation (Wu et al. 2007). The Wright Map for the Arctic Trek data yielded similar results to the map in the Webspiration case.

In summary, these preliminary results show that it is indeed feasible to collect data on a new variable such as Learning in Networks, and to do so using innovative item types that encompass web resources. The reliability coefficients that were observed are quite strong, even though the number of items in Arctic Trek was not very large. The good match between the expected levels of response and the empirical results indicates quite sound levels of internal structure validity.

Conclusion and Next Steps

Measuring collaborative digital literacy as described here is helping us understand how students think and work differently than in previous decades. Accessing, using, and creating information and knowledge digitally employs many important skills needed today for career and college readiness. This chapter describes a domain modelling process for assessing these skills through the BEAR assessment system, along with examples of task development, and results from implementation of a pilot study in four countries.

However, the domain modelling process is as yet incomplete for this set of constructs. The hypothesis indicated in Fig. 3.1 has not yet been properly tested (that will need to wait until we have a larger data set from field trials) and, indeed, the final form of the hypothesised structure is also incomplete. What is as yet missing is a next level of elaboration of the learning progression, which is characterised by hypothesised links between the levels of different constructs. The substantive and empirical discovery process that establishes these hypotheses is not yet complete, but the full diagram will be more like the one shown in Fig. 3.20. This learning progression is from a separate project, the Assessing Data Modeling project (Lehrer et al. 2014). In this project, there are seven constructs, shown here as the vertical sets of blocks (each block representing a level of the construct). Between some levels of some constructs are arrows, which indicate hypothesised hierarchical links between those levels. The probability of students being observed in the target level (i.e., the level the arrow points to) is expected to be very low, unless they have already shown evidence of being at the source level (i.e., at the other end of the arrow). This presentation allows the incorporation of interesting educational information about how students are expected to progress through the skills and knowledge defined in through the learning progression. A hypothetical learning progression for Learning in Networks is shown in Fig. 3.21. Statistical models to estimate these links are currently being developed (Wilson 2012) and will be available for use when field test data is collected.

The participating ATC21S countries through the first phases of the project have helped illustrate how their teachers and school systems support students to develop twenty-first century competencies. Conclusions from the pilot studies show that students in the 11–15 year age group demonstrate widely differing knowledge and skills in these areas. Some are only beginning to take their first tentative steps toward digital competence while others exhibit quite breathtaking levels of mastery, such as the ability to collaborate seamlessly to create in mere moments insightful audio commentaries and share them for common understanding. Differences in what

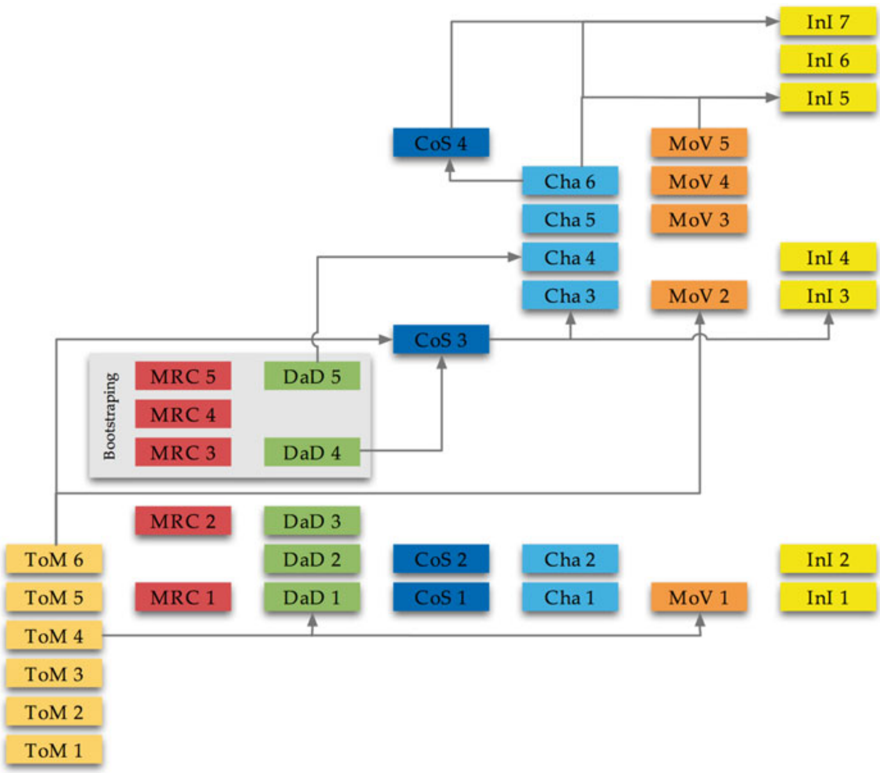


Fig. 3.20 An example learning progression diagram from the ADM project

students can do, and the absence of formal teaching and opportunities to learn these skills, point to a rapidly widening gap between important LDN-ICT skills and what schools offer. ATC21S results are showing this to be particularly true when collaboration, creation and problem-solving are involved, based on such early assessment efforts as described here.

The next steps for ATC21S involve wide-scale fieldwork trials for a segment of the tasks, currently drawn from the collaborative problem-solving domain, now being conducted in Australia, Finland, Singapore and the U.S.A. Associate countries Costa Rica and the Netherlands are joining in to help test how language and culture affect twenty-first-century teaching and assessments. The digital literacy domain tasks described here are being used to explore the language and culture localisation process.

The final phase of the project will place the ATC21S resources in the public domain. This will allow government policy-makers, teachers, school systems and assessment institutions to download, modify and extend existing research and materials. This may help to bring more broadly the twenty-first-century skill domains described here into classrooms around the world. Certainly an important contribution is to encourage more conversation on how information-age trends do not stop at the school door.

Level	CiN Description
C	Lorem ipsum dolor sit amet, consectetur adipiscing elit.
B	Cras males uada tincidunt tincidunt. Cras facilisis fringilla neque nec auctor.
A	Maecenas euismod duinon leo pretium.

Level	PiN Description
C	Nam at nulla arcu. Pellentesque fermentum enim turpis.
B	Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nullam turpis metus.
A	Praesent interdum magna vitae est posuere, non faucibus dolor auctor. Proin tempor nisi quis leo adipiscing.

Level	SCN Description
D	Donec aliquet pulvinar est. Nullam scelerisque risus sed est porttitor.
C	Vestibulum laoreet dui nisi, quis posuere odio faucibus nec. Nullam lacinia tempus porttitor.
B	Lorem ipsum dolor sit amet, consectetur adipiscing elit. Donec ultricies enim.
A	Pellentesque habitant morbi tristique senectus et netus et malesuada.

Level	ICN Description
D	Praesent interdum magna vitae est posuere, non faucibus.
C	Vivamus vel nibh felis. Vivamus pharetra urna sollicitudin pretium porta.
B	Pellentesque acfringilla est. Nunc semper nunc ligula, consequat auctor ligula feugiat sit amet.
A	Pellentesque ac fringilla est. Nunc semper nunc ligula, consequat auctor ligula feugiat sit amet.

Fig. 3.21 A hypothetical learning progression for learning in digital networks

Acknowledgements We thank the ATC21S project and its funders for their support for the work reported here. We also acknowledge the expertise and creative input of the ATC21S Expert Panel in LDN-ICT: John Ainley (Chair), Julian Fraillon, Peter Pirolli, Jean-Paul Reeff, Kathleen Scalise and Mark Wilson. Of course, the views and opinions expressed in this paper are those of the authors alone.

References

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*. Dordrecht: Springer.

Cisco, Intel, Microsoft (CIM). (2008). *Transforming Education: Assessing and Teaching 21st Century Skills.* Authors. Downloaded from: http://atc21s.org/wp-content/uploads/2011/04/Cisco-Intel-Microsoft-Assessment-Call-to-Action.pdf

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

Lehrer, R., Kim, M.-J., Ayers, E., & Wilson, M. (2014). Toward establishing a learning progression to support the development of statistical reasoning. In J. Confrey & A. Maloney (Eds.), *Learning over time: Learning trajectories in mathematics education*. Charlotte, NC: Information Age Publishers.

Masters, G. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149–174.

National Research Council (NRC). (2001). In Committee on the Foundations of Assessment, J. Pellegrino, N. Chudowsky, & R. Glaser (Eds.), *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.

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

Wilson, M. (2009a, December). *Assessment for learning* and *for accountability*. Paper presented at the exploratory seminar: Measurement challenges within the race to the top agenda, at ETS, Princeton, NJ.

Wilson, M. (2009b, December). *Assessment for learning* and *for accountability*. Policy brief from the exploratory seminar: Measurement challenges within the race to the top agenda, at ETS, Princeton, NJ.

Wilson, M. (2012). Responding to a challenge that learning progressions pose to measurement practice: Hypothesized links between dimensions of the outcome progression. In A. C. Alonzo & A. W. Gotwals (Eds.), *Learning progressions in science*. Rotterdam: Sense Publishers.

Wilson, M., & Sloane, K. (2000). From principles to practice: An embedded assessment system. *Applied Measurement in Education, 13*(2), 181–208. Download from: http://www.informaworld.com/smpp/content~content=a783685281~db=all

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*. Dordrecht: Springer.

Wu, M., Adams, R., Wilson, M., & Haldane, S. (2007). *ConQuest: Generalised item response modelling software (Version 2.0)*. Camberwell: ACER Press.

Part III
Delivery of Collaborative Tasks, Their Scoring, Calibration and Interpretation

In Chap. 4, Care, Griffin, Scoular, Awwal and Zoanetti (2015) describe the prototype collaborative problem solving tasks. The tasks were constructed as games through which students collaborate to solve problems or learn. The tasks can be divided into two types. There are those that are curriculum independent and those that are curriculum dependent. In addition there are those that are symmetric and those that are asymmetric. There are tasks that are presented as a single internet web page and tasks that are multipage. Multipage tasks were designed to become increasingly difficult and complex with increasing numbers of pages. In Chap. 5, Awwal, Griffin and Scalise (2015) describe the delivery platform which houses the tasks and controls access, security and data collection. The links between the task bank and the collaborative allocation enables contributions of individual students to be determined. In Chap. 6, Adams, Vista, Scoular, Awwal, Griffin and Care (2015) describe how the data is collected. They also describe how the platform is used to apply the coding and scoring algorithms to produce the student performance reports for teachers. Griffin, Care and Harding (2015, Chap. 7) demonstrate how the data is interpreted and how scores are calibrated using item response modelling, and they present the dimensions of the domains. The chapter presents evidence of the construct validity, stability of indicators across systems of education, across curricula, across languages, and provides evidence that the construct being measured across those contexts is a constant. Details of tasks are provided in terms of the structure, symmetry and complexity and an increasing shift towards human to human interaction on the Internet.

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

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

Adams, R., Vista, A., Scoular, C., Awwal, N., Griffin, P., & Care, E. (2015). Automatic coding procedures for collaborative problem solving. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills: Methods and approach* (pp. 115–132). Dordrecht: Springer.

Griffin, P., Care, E., & Harding, S. (2015). Task characteristics and calibration. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills: Methods and approach* (pp. 113–178). Dordrecht: Springer.

Chapter 4
Collaborative Problem Solving Tasks

Esther Care, Patrick Griffin, Claire Scoular, Nafisa Awwal, and Nathan Zoanetti

Abstract This chapter outlines two distinct types of collaborative problem solving tasks – content-free and content-dependent – each allowing students to apply different strategies to solve problems collaboratively. Content-free tasks were developed to emphasise the enhancement of inductive and deductive thinking skills. Content-dependent tasks allow students to draw on knowledge gained through traditional learning areas or subjects within the curriculum. The collaborative problem solving framework emphasises communication for the purpose of information gathering, identification of available and required information, identification and analysis of patterns in the data, formulation of contingencies or rules, generalisation of rules, and test hypotheses. Characteristics of tasks which were identified as appropriate for eliciting collaborative problem solving processes are reported and illustrated by exemplar items.

Introduction

This chapter demonstrates how the collaborative problem solving (CPS) framework, outlined in Hesse et al. (2015; Chap. 2), is applied to a selection of tasks and, in turn, how each of the tasks highlights the skills outlined in the framework. There are two distinct types of tasks presented here: content-free and content-dependent. Content-free tasks do not demand any prerequisite knowledge such as might be taught in traditional school-based subjects but rely on the application of reasoning. Content-dependent tasks draw on skills and knowledge derived from

The views expressed N. Zoanetti, in this chapter are those of the author and do not necessarily reflect the views of the Victorian Curriculum and Assessment Authority.

E. Care (✉) • P. Griffin • C. Scoular • N. Awwal
Assessment Research Centre, Melbourne Graduate School of Education, University of Melbourne, Parkville, VIC, Australia
e-mail: e.care@unimelb.edu.au

N. Zoanetti
Victorian Curriculum and Assessment Authority, East Melbourne, VIC, Australia

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_4

curriculum-based work. As discussed in Hesse et al. (2015), under the proposed CPS framework there are three strands of indicators that summarise social skills and reflect the collaborative aspect of problem solving: participation, perspective taking, and social regulation. Participation is the foundation for engaging with the task and other collaborators, and is reflected in the way people act or interact to complete tasks. Perspective taking skills emphasise the quality of interaction between students, reflecting the level of student's awareness of their collaborators' knowledge and resources as well as their responding skills. Social regulation refers to the strategies used by students when collaborating, such as negotiating, taking initiative, self-evaluating and taking responsibility. Cognitive skills are of equal importance within this framework and are similar to those employed in independent problem solving tasks. Indicators of such skills can be summarised under two headings: task regulation and knowledge building. Task regulation refers to the ability of students to set goals, manage resources, analyse and organise the problem space, explore a problem systematically, aggregate information and tolerate ambiguity. Knowledge building is concerned with a student's ability to understand the problem and to test hypotheses. Knowledge building is underpinned by skills such as planning and executing, and reflecting and monitoring.

In teaching students how to become better problem solvers, a common constraint in traditional test design has been that the attainment of the solution is the sole criterion from which inferences can be made. This has occurred despite the fact that procedural aspects of problem solving have been considered important for some time (Polya 1945, 1957; Garofalo and Lester 1985; Schoenfeld 1985). Within the ATC21S project[1] there is an increased focus on drawing inferences about how (and how well) students solve problems, as opposed to simply asking whether they are solving them. Problem solving has sequential phases or steps, such as understanding, planning, solving and checking, that are universally applicable across tasks and contexts. This information, together with information on student collaborative effort, might better support the decisions an educator must make when determining the instructional needs of individual students (Zoanetti 2010). Although goal-attainment is obviously important, it should not be the only criterion of interest. Educators stand to benefit from inferences about procedural quality when determining how best to improve student problem solving.

Problem and Task Characteristics

The differences between real-world problems and problems as they are often analysed in psychological research raise the question of whether the assessment of collaborative problem solving through well-defined problems is useful. A "well-defined" problem is one in which the guiding question and consequently the

[1] The acronym ATC21S™ has been globally trademarked. For purposes of simplicity the acronym is presented throughout the chapter as ATC21S.

goal is known, where the elements or "artefacts" that are salient to the solution are known and present, and where the required processes to reach solution are understood. Such problems are amenable to measurement since they involve specific known steps, and have final correct solutions. Use of these types of problems also lend themselves to teaching since a sequence of steps is often clear. Well-defined tasks are typically found within the science and mathematics curriculum. On the other hand, "ill-defined" problems are characterised by ambiguity. They may relate to everyday problems and are not domain-specific; they may draw on many different types of knowledge. They will have many of the characteristics that are associated with what is known as "wicked" problems. These are problems in the real sense of the word – situations for which a solution is unknown, of which the elements or components are not identified, and concerning which useful processes have not been verified. Consequently, for ill-defined tasks there may be several solutions that are appropriate to different degrees, several solution paths or strategies, and it may be the case that not all information is presented or available. There may be no clear direction in which to proceed and no clear identification of how the correctness of a solution can be determined.

The difference between well-defined and ill-defined problems calls into question how valid might be the inferences about individuals' problem solving capacities if drawn only from well-defined problems. The long term objective of teaching problem solving skills would be to equip students with the capacity to draw from a range of strategies when confronted with ill-defined problems – which latter actually constitute the real-world imperative.

Hesse et al. (2015) describe the nature of problems that might require collaborative activity. The salient feature is that resources will not be equally accessible to all the problem solvers, so there is a need for multiple solvers. Accessibility refers both to direct retrieval as well as to human capacity to understand and manipulate the required artefacts – whether these be objects, knowledge, or processes.

Together, the concerns about whether only well-defined problems can usefully indicate students' problem solving capabilities, and the nature of problems that require collaborative activity, combined within the ATC21S approach to the deliberate design of tasks along a well-defined to ill-defined spectrum. The assessment tasks were constructed to reflect the characteristics of problems which require collaboration. These characteristics are ambiguity, asymmetry, and unique access to resources with consequent dependence between learners. With such tasks it is possible to test the construct definition model, the developmental learning progressions, the indicators of increasing competence, and the task development and delivery. At the most simple level, problem solving tasks were designed to make collaboration both desirable and essential. In the classroom, this can be achieved by the teacher giving different sets of information to different students in a group, rather than giving them all the same information. In order to solve the problem, the students then need to collaborate in order to access the required resource, in this case, information. Such an approach mirrors real life collaborative problem solving situations, where information may be derived from different sources and is not shared a priori. The dependence between learners that emanates from unique access to different resources provides a more authentic prompt for collaborative activity than mere instructions from a teacher for students to “work together”. Working together may be valued for its social aspect, yet might not be essential, and can be regarded by students as counter to their best interests – particularly when they are functioning in competitive classroom environments.

The tasks in the ATC21S project have many similar characteristics. Each task was constructed so that students would be able to click, drag and drop objects using the mouse cursor, with no requirement to use the keyboard. The tasks were designed for two students to work on and there is a ‘chat box’ for communication between collaborators, designed to facilitate student communication online throughout task completion. Each task presents an instruction stem followed by a problem with tasks ranging from 1 to 8 pages in length. The tasks were designed to be recognisable at face value as puzzles and to include graphics to attract and maintain student engagement. A few of the tasks present exactly the same images, perspectives, instructions and resources to the two students – these are referred to as symmetrical tasks. Many of the tasks present asymmetrical perspectives, providing different information and resources to each student, thereby increasing their need for collaboration. There is encouragement in the tasks for students to discuss the problem in order to manage the identification of resources, and sharing of these. The tasks vary in difficulty level; some require less collaboration but are cognitively more difficult, while others are cognitively easier but require efficient collaboration to solve. The difficulty of the tasks was varied taking into consideration arguments of Funke (1991) by adjusting several of the parameters, such as the number of problem states, the constraints on object manipulation built into each task and described in the problem stem, the complexity of reasoning or planning required to guide the search, and finally the configuration of objects and symmetries within the task.

The matter of symmetry poses challenges to assumptions made in education about equal access for learners. Although there may well be major differences in education provision across and within countries, the presumption is that in any classroom all students will have the same access to resources. In this context, resources refer to tools, texts, teachers, and the classroom environment with all of these supporting and enhancing the learning of the student. This provision is extended to equality of access in the assessment situation, with all students again typically being provided with the same resources. This equality of access has been contested in the last decade by virtue of emphasis in some learning environments, on group work. In this scenario, equality of resource is not assured, since different groups will present with different human resources, and the capacity of the individual to act will be determined not only by their access to resources, and their own capacities, but also by the capacities of others. This reality is reflected in the ATC21S assessment environment, where students are not provided with the same access to resources – either those constructed within the assessment environment, or those that ensue from the varying capacities that student partners bring into play. Both differential access to resources and the consequent dependence between students bring about asymmetry in the assessment task activity.

Asymmetry raises interesting challenges in the world of assessment, as well as in how students and their teachers cope with the learning and teaching activity. In this chapter we demonstrate how both symmetry and asymmetry is manifested in the assessment environment. Discussion of the consequences of this for scoring is presented in Adams et al. (2015; Chap. 6).

Content-Free Collaborative Problem Solving Tasks

Two tasks outlined in this section focus on students' hypothetico-deductive reasoning skills in an online collaborative problem solving context. The translation of these steps into a process that can be generalised and called "collaborative problem solving" should enable teachers to assess and develop their students' capacity for hypothetico-deductive thinking as it manifests itself in collaborative problem solving behaviour. Hypothetico-deductive thinking begins with a causal question. Students then generate hypotheses based on observations and data collection. In a virtual world it is possible to monitor this behaviour through analysis of chat and action events. These events can be seen to follow a pattern suggested by Griffin (2014), who argued that problem solving can be understood as a hierarchical series of steps moving from inductive to deductive thinking. Problem solvers first examine the problem space to identify its elements. Next they recognise patterns and relationships between the elements, and formulate these into rules. The rules are then generalised. When generalisations are tested for alternative outcomes, the problem solver is said to be testing hypotheses. 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. The deductive method attempts to "deduce" facts by eliminating all possible outcomes that do not fit the available information. Collaborative problem solving requires the formation of partnerships in which agreement is reached on the nature of hypotheses to be tested and the manner in which they will be tested.

The two "content-free" tasks described here are compatible with an individual problem solving approach in that each has a finite solution, and all the information required for problem solution is included in the problem space. The transition to identification of these tasks as collaborative problem solving tasks lies in the re-structuring of the problem space such that neither member of a pair of collaborating students has access to all necessary information. The first task, Laughing Clowns, is structured symmetrically – both students have access to all resources; while the second task, Olive Oil, is structured asymmetrically – each student has access to different resources. The term "problem space" here refers to the virtual environment which provides all the stimuli and resources that identify that there is a problem. The stimuli include text instructions and some explanation about the problem, as well as virtual artefacts, both static and dynamic, including the graphic objects on the screens, and the indicators of movement such as mouse cursor.

The tasks are hosted on a virtual platform that allows for real-time work activity by two students operating in a one-to-one computing environment. Students may work on the tasks on any computers that have internet access and up to date browsers. Technical requirements are outlined by Awwal et al. (2015; Chap. 5). Each task is described here in terms of the problem solving goals, and the activities or processes and artefacts available to the students. The description is followed by an analysis of the subskills from the conceptual framework that are drawn upon, and assessed through the task.

Laughing Clowns Task

This task requires students to find patterns, share resources, form rules and reach conclusions. The two students are presented with a clown machine and 12 balls to be shared between them. The goal for the students is to determine whether their clown machines work in the same way. In order to do this, the two 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 the clown's mouth while it is moving in order to determine the rule governing the direction the balls will go (Entry = Left, Middle, Right, and Exit = position 1, 2, 3). Each student must then indicate whether or not they believe the two machines work in the same way (see Fig. 4.1). Students do not have access to each other's screen so are not able to determine the rule governing the other's clown machine.

Social Skill: Interaction

A fundamental requirement for successful completion of this task is interaction between partners. Students need to be aware from the start that their 12 allocated balls are shared and that the most effective way of finding the solution is to allocate six balls to each such that both students have adequate and equal opportunity to trial their machine and reach a conclusion. Students who do not interact may begin using the balls, and even use them all before realising the resources are shared. More

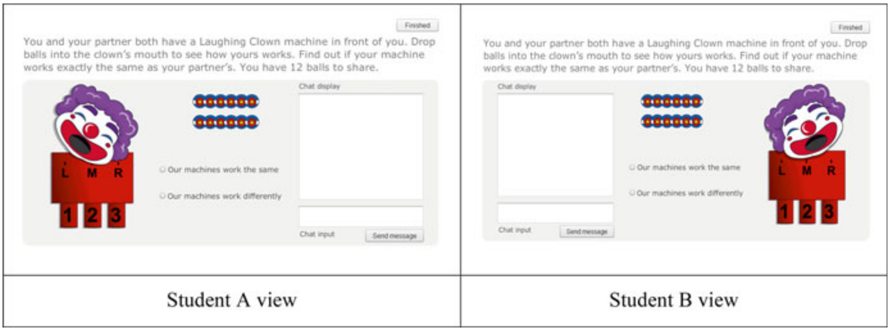


Fig. 4.1 Laughing Clowns task

proficient students are likely to be aware early in the task of the need to coordinate their and their partner's activity and will promote interaction with their partner before they begin to use the balls and test their own machine.

Social Skill: Audience Awareness

Students possessing good perspective taking skills would be aware of their partner's role in this task and the need to understand their partner's perspective. Students who do not possess strong skills in this area are likely to proceed with the task with little consideration for their partner's resource requirements or observations. Students who are proficient are likely to interact with their partner in between ball drops and adapt their behaviour to best suit their partner's needs. An indicator of this skill is the number of moves students make before stopping and waiting for their partner to move or respond, fewer moves being, in this case, the preferred response.

Cognitive Skill: Resource Management

The ability to manage the available resources contributes to a student's ability to regulate the task well. For example, students who have lower proficiency in this skill may only concern themselves with checking on how their own machine functions, thereby monopolising use of the resources, while more proficient students are likely to recognise the need for shared use of the balls and allocate them equally.

Cognitive Skill: Relationships (Representing and Formulating)

Students must identify the relationship between entry and exit point of balls, and determine if there is a consistency in how the machine functions. They then need to construct a way of representing this information that will communicate to the partner, as well as being able to understand other forms of representation that the partner uses. The student may choose to represent the relationships through listing discrete pieces of information, through narrative, or through formulation of rules. Each of these representations needs to be amenable to communication through the chat box which is part of every task. Proficient students will also challenge patterns and test the assumptions that underpin their observations – for example, consistency of patterns. The final step comprises the students comparing their representations such that a decision concerning similarity of clown machine functioning can be made (Table 4.1).

Olive Oil Task

In this task students are presented with different resources. In order to achieve the objective of the task – which is to fill a jar with 4 l of olive oil – the students must work out what resources are available and are needed. Student A has a virtual three

Table 4.1 Example of skills observed in Laughing Clown task

Skill	Behaviour	An example of data captured for assessing
Interaction	Interacting with partner	Presence of chat before allowing partner to make a move
Audience awareness	Adapts contributions to increase understanding for partner	Number of ball moves attempted before stopping and waiting for partner to move or respond
Responsibility initiative	Takes responsibility for progress for the group task	Number of times communicated with partner before the first half of the shared balls is used up
Resource management	Manages resources	Realises that balls are meant to be shared and uses only half of the available
Systematicity	Implements possible solutions to a problem	Uses half of the balls to cover the positions in a sequential order
Relationships	Identifies connections and patterns between elements of knowledge	The two students come to an agreement on how their machine works
Solution	Correct answer	Selection of the correct option by Students A and B on how their machines work

litre jar, olive oil dispenser, transfer pipe and bucket. Student B has a virtual 5 l jar, transfer pipe and bucket. Without knowing what is available to the other, the pair need to recognise that Student A must fill their jar at the dispenser and place it under the transfer pipe so that Student B can accept the oil from the pipe. Until this point, Student B cannot complete any meaningful actions and is dependent on the actions and interactions of Student A. Students need to explore and navigate the task space together until they can place 4 l in Student B's jar. This task follows the reasoning processes required in the Tower of Hanoi problem popularised by mathematician Eduard Lucas in 1883 (Newell and Simon 1972; Petković 2009). The problem requires the solver to work out a sequence of movements to achieve the goal. It bears some resemblance to the forward planning requirements of a chess game – in thinking beyond one step to the next before implementing an action. This cognitive task is made more complex by the division of resources and the lack of information available to each student (see Fig. 4.2).

Social Skill: Interaction

While current technologies do not afford us the ability to analyse the actual text of the communication, the quality of interaction can be inferred through the placement of chat. In this task, interaction is assessed by the presence of chat during specific problem solving stages or 'blocks' indicating the level of interaction between students and the perceived importance of collaboration during specific processes.

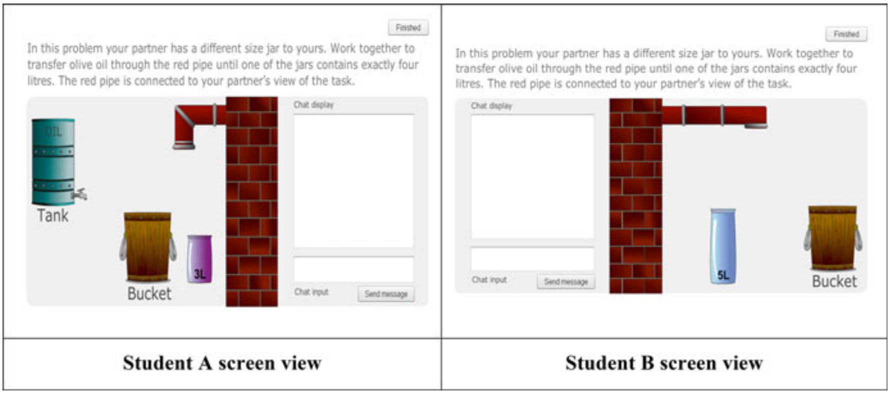


Fig. 4.2 Olive oil task

Students who have strong communication skills may initiate or prompt the interaction immediately after the task begins.

Cognitive Skill: Cause and Effect

The ability to use their understanding of cause and effect to develop a plan will enhance a student's success in this task. A way to measure planning and executing skills is to assess the amount of time taken between actions. For example, after realising that their jar is empty at the start of the task, more able students operating as Student A will take a shorter amount of time than less able students to fill the jar at the dispenser. Another indicator of successful planning and executing for Student A is the time taken between their jar containing 1 l of oil and the transfer of that litre to Student B. This requires students to think of steps ahead of their current state and work out sub-tasks before acting. Some students may propose several rules of cause and effect before gaining success. An example is the presence of the bucket for Student A. This object is redundant but Student A may use the bucket to empty the 3 l jar before realising that this action does not provide a pathway to problem solution.

Cognitive Skill: Problem Analysis

Proficient students are able to analyse the problem before organising the necessary steps to solve it. One example of a student analysing the problem is the identification of their need for information and resources from their partner – which requires elements of task regulation – followed by their description of this problem in a mode of communication familiar to their partner. An indicator of this is the exchange of information, assessed by the presence of the key numbers (1, 3) within the chat,

Table 4.2 Example of skills observed in the olive oil task

Skill	Behaviour	An example of data captured for assessing
Interaction	Interacting with partner	Presence of chat during a specific set of actions and processes
Cause and effect	Identifies sequence of cause and effect	When A's 3 L contains only 1 L, A recognises that this must be transferred to B
Reflects and monitors	Adapts reasoning or course of action as information or circumstances change	Learning from redundant activities, such as A moving jar to bucket
Relationships	Identifies connections and patterns between and among elements of knowledge	Presence of chat exchanging information when A or B recognises significance of their jar containing only 1 L
Solution	Correct answer	Last action requires B's jar to contain 4 L of oil
Problem analysis	Identifies necessary sequence of subtasks	Exchange of important information during necessary sequence

during the time that Student B's jar contains 1 l of oil, followed by the acceptance by Student B of the 3 l of oil.

Cognitive Skill: Solution

Although students' proficiencies are not being measured predominantly on their success or failure in completing the task, this factor is still measured. In this task, we assess whether students found the correct solution to the problem by checking whether their final action results in Student B's jar containing 4 l. The steps taken to solve the problem can then be assessed to determine the processes used and the students' efficiency in achieving the solution (Table 4.2).

Content-Dependent Collaborative Problem Solving Tasks

The content-dependent tasks draw on particular skills and knowledge derived from school or curriculum based work. These tasks stimulate the development of assessable curriculum-linked problems that can be solved collaboratively and that connect with everyday teaching and learning in the mathematics and science curricula around the world. In the examples presented here only basic subject based knowledge is required.

The two content-dependent tasks outlined here were originally designed by World Class Arena Limited (WCAL) for use as online single student problem solving tasks. Under contract with WCAL the tasks were redesigned for use as collaborative tasks. This involved redesigning tasks so that they required iterations between collaborators, not merely a division of labour. The tasks were designed to be complex, unscaffolded and ill defined. The lack of scaffolding lies in omitting guidelines for the students that would help them to understand both how to proceed, and the fact that there might be multiple paths that could be followed. In the initial design of the single student tasks the problem solution was much simpler to achieve. Students could easily follow the path to solution by understanding the problem, selecting a strategy, and applying the strategy. The path to solution was a simple one and collaboration within this context would not provide much in the way of additional support, information, ideas or resources. To redesign the tasks to be more complex, the stages to problem solution needed to be less clear. This stimulated more sophisticated strategies that require both collaborators to be active participants. Together students are required to try several different strategies to solve the problem, sharing information with one another and reflecting before trying an alternative solution path. The collaboration between problem solvers is a parallel rather than a serial process. It is anticipated that students will be able to better understand the problem using this rigorous method of investigation and develop the ability to transfer this knowledge successfully to different contextual scenarios. Optimally, each collaborator is fully involved in each stage of the process, such that both will reach and agree on the problem solution, and gain an understanding of the process.

Within each of these tasks the complexity increases through subsequent pages with varying approaches to the problem, and allowing knowledge to build. Each page requires both students to participate in the task, and only together can either proceed to the subsequent page of the task. In this respect one student's progress is linked to the other. It was the intention that this level of scaffolding would prompt students to communicate. Generally, with the content-dependent tasks the final page is designed for independent working and therefore both students must have gained enough knowledge collaboratively in order to apply their knowledge to the final answer independently. Communication on the final pages of the tasks is encouraged in order to optimize the chance that each partner has fully understood the problem, since their task completion jointly depends on it. Although complexity increases throughout the pages, if sufficient knowledge building and task regulation have taken place on previous pages, the final subtask should not present greater difficulty than experienced earlier in the task.

Balance Beam Task

The Balance Beam task is an example of a content-dependent task with elements somewhat reliant on an understanding of the science – in this case, the physics – behind the task. Students need to apply and test rules in order to balance the beam, leading to multiple correct solutions. Collaborating students share a balance beam but each can interact with only one side of the beam (see Fig. 4.3). Student A begins with four different masses, Student B has no masses, and is not directly aware of

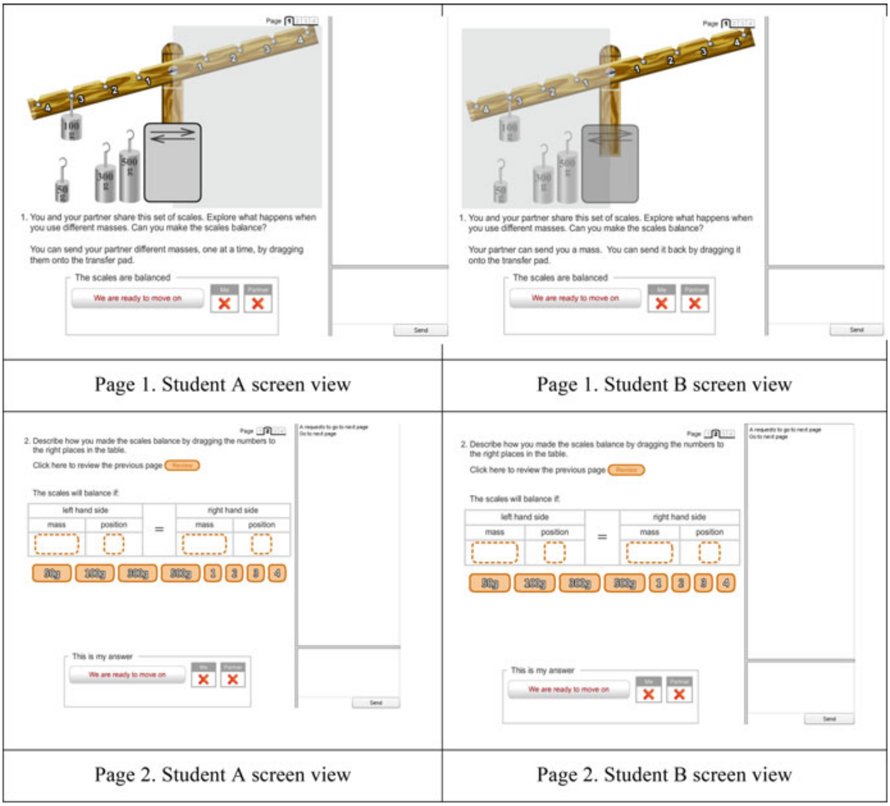


Fig. 4.3 Pages 1 and 2 of the balance task

this resource. Initially masses must be passed from Student A to Student B, and each must place their masses in the correct notch on the beam in order to achieve balance. The students are able to pass the masses back and forth to each other. There are four notches on each side of the balance beam.

On the second page of the task students are asked to provide the formula which best describes how they balanced the beam. In subsequent pages students are required to balance the beam in several different ways. The additional pages were constructed to ensure that the students understand the physics of the problem and reduce the probability of successful guessing. Examples are provided below of the elements within the theoretical framework that underpin the social and cognitive processes required to complete the task.

Social Skill: Responsiveness

The student needs to adapt and incorporate contributions from the other. One way in which this can be evaluated is to assess which masses have been transferred by Student A to Student B. If the correct masses are sent, we can infer that the student has grasped the task concept. A specific indicator measures whether Student A sends particular masses to Student B and whether Student B returns them immediately. The latter identifies that Student B successfully responds to their partner by acknowledging which resources are the most useful, and permits the inference that the student understands the task concept.

Cognitive Skill: Systematicity

Within this task it is possible to measure how systematically a student approaches the task. Where a specific sequence of actions is identified, it can be assessed to determine how students are implementing possible solutions and monitoring their progress. For example, functional systematicity skills can be assessed by measuring the number of trials of balance attempted by the collaborators. This can be done by counting the positions tried for each and all masses. Too few or too many would suggest a lack of systematicity. If a student tests every position once and exhausts all possible combinations, they are exploring the space and the resources thoroughly. An example of poor systematicity within this task is Student A continually attempting to pass a further mass to Student B when B already has the maximum number of masses permitted. Contingent on how the task is designed, this may mean that the student is not approaching the task systematically or monitoring their own actions efficiently. They may not have understood the task instructions or not identified the structure of the task (Vollmeyer et al. 2006), or they may not be learning from their mistakes.

Cognitive Skill: Sets Goals

Goal setting is a key skill in problem solving and can be measured in various ways across the assessment tasks. One example within the Balance Beam task is the presence of a numerical value within the chat which represents one of the mass amounts (100 g, 200 g, etc.). If the chat is from the student who does not have those corresponding mass amounts it can be inferred that the student is requesting those mass amounts from their partner and that their goal is to use them to balance the beam. In addition, if the mass amounts are the correct ones (that is, they would balance the beam) it can be inferred that the student has understood the physics underlying the task and intends to use the masses to attempt problem solution (Table 4.3).

Game of 20 Task

The Game of 20 task involves students working together against the computer to reach a value of 20 by placing counters sequentially on a grid. The students need to identify crucial scores and limits in order to win the game. This task involves algebra equations relevant to the mathematics curriculum. There are six pages within the

Table 4.3 Example of skills observed in the balance beam task

Skill	Behaviour	An example of data captured for assessing
Action	Active in scaffolded environments	Student A passes B a mass
Task completion	Undertaking part of a task individually	Follows instructions, moves 100 g to position 4
Responsiveness	Responding to contributions of others	Realises that some masses cannot balance. If Student A resends 50 or 500, B returns it immediately
Sets goals	Sets goals for a task	Requests mass amounts
Systematicity	Implements possible solutions to a problem	Trial of different combinations of masses on different beam positions
Solution	Correct answer	Number of successful balances achieved (3 optimum)

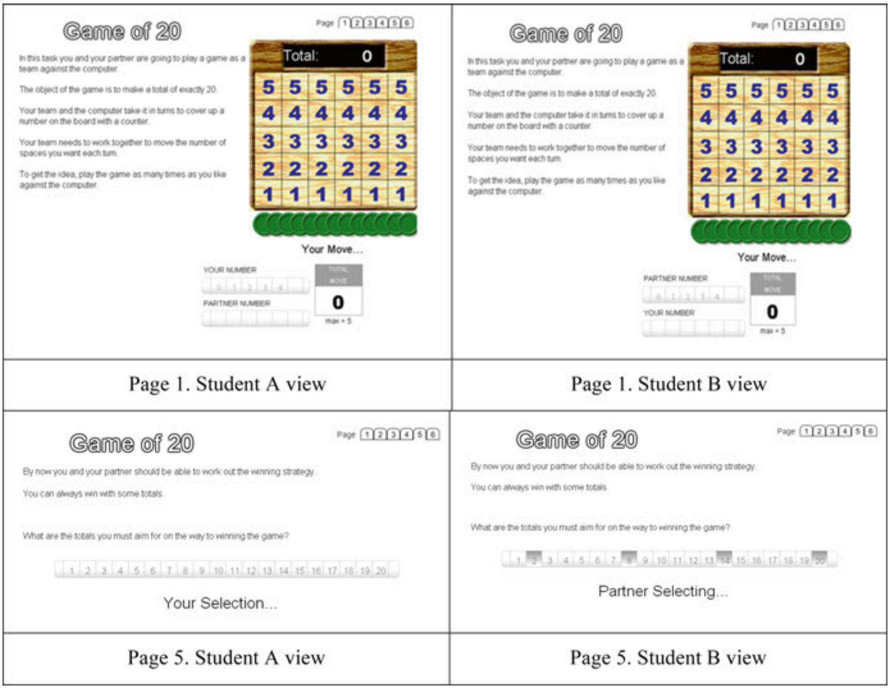


Fig. 4.4 Pages 1 and 5 of game of 20 task

task; five of them present a sub-task where students begin the game at various stages, thereby allowing for varying degrees of difficulty. The first page allows students to play the game in full to help their understanding of the task concept (see Fig. 4.4). On the second page students begin with a game total of 18; on the third page they begin with a game total of 13, and on the fourth page with a game total of six. Page 5 presents students with a number line from which together they select the numbers they believe to be crucial to success in the game. Students are assessed on whether they implement the numbers selected within game play on the subsequent task page (page 6) when they replay the game in full.

In order to play the game, students independently choose a number between zero and four which contributes to the combined team number from one to five (that is, the two student values are combined to create their team number). Each student needs to consider the input of their partner when selecting a number, as well as considering which number the computer will select. There are several rounds of number selection until the game total reaches 20. If either team (students or computer) enters an amount that exceeds the game total of 20, then that team loses. The aim of the game is for the student team to reach the exact game total of 20 before the computer does.

Social Skill: Responsiveness

Within the Game of 20 task, students' ability to ignore, accept or adapt contributions from their partner can be assessed. Students who are strong in this skill may be observed selecting a specific number after their partner has sent them a chat message containing that number. It can be inferred from this indicator that their partner has contributed to their activity and that the student has accepted and incorporated this contribution into their game play. Another example of this skill may be observed as students work through page 5, where they need to agree on which numbers are crucial for game success by selecting them on the number line. Students who are less adept in this skill may not accept or consider contributions from their partner even if they are correct. This can be observed when a student deselects a number from the line that their partner has previously selected.

Social Skill: Responsibility Initiative

Students who are more collaborative tend to take more responsibility for their team and ensure that the activities required for task success are completed by themselves and their partner. One example within the Game of 20 task is a student attempting an activity and then reporting their actions to their partner. This may be observed if a student resets the team number, chats with their partner and then changes the numbers selected before progressing with the game. It can be inferred from this activity that the student was not satisfied with the initial number selection, opted to reset it, and then reported this to their partner, resulting in alternative number options.

Cognitive Skill: Cause and Effect

The extent to which students use their understanding of cause and effect to plan and execute can be assessed in this task. Students who are less proficient in this skill may undertake the activity with no clear regard for the consequences of their actions, but more proficient students will use their understanding of cause and effect to plan and execute a strategy or activity. An example of a proficient student in this task is one who selects specific numbers on the number line and then proceeds to use those numbers in game play on subsequent task pages. It can be inferred from this action that the student has determined from previous game play that these numbers are crucial to success and intends to use them in order to succeed in the game.

Cognitive Skill: Reflects and Monitors (Testing Hypothesis)

Students can be assessed on their ability to hypothesise effectively. Students who are not effective in formulating hypotheses tend to maintain one single approach throughout a task, are not flexible and therefore fail to monitor their progress efficiently. Students who have strong skills in developing hypotheses tend to reflect more on their previous actions, monitor their progress, reorganize a problem and try multiple approaches as they gain further information. An example within this task is a student who opts to retry the game after they have already attempted it. On each task page the student can opt to retry each sub task page. By doing so the student is attempting to reflect on the course of action that caused them to fail previously and is trying another approach in order to gain a different outcome and a successful solution (Table 4.4).

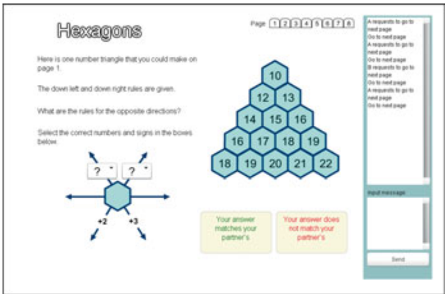
Table 4.4 Example of skills observed in the game of 20 task

Skill	Behaviour	An example of data captured for assessing
Task completion	Undertaking and completing part of a task	There's a win before moving on.
Responsiveness	Responding to contributions of others	Responsiveness to chat from partner containing crucial number
Responsibility initiative	Takes responsibility for progress of the group task	Alters plan and exchanges information that suggests further planning
Cause and effect	Identifies sequence of cause and effect	After sufficient game play , selects numbers on number line that are crucial to success
Reflects and monitors	Adapts reasoning or course of action as information or circumstances change	Replaying the task after a failed attempt
Collects information	Collects information	Presence and count of questions in the chat

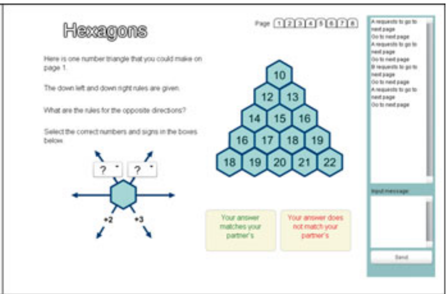
Conclusion
These task descriptions link problem solving and collaborative activities required of students as they engage with tasks. The tasks are engineered to provide opportunities for demonstrations of skills hypothesised to contribute to collaborative problem solving capacities. This approach to task construction reflects clearly the use of inference to attribute meaning to student test responses. Where we are interested in skills development and progression, as opposed to degree of finite content, skills, or knowledge held by individuals, such inferential approaches are essential.

Appendix: Collaborative Problem Solving Tasks

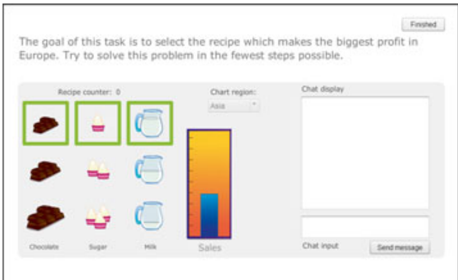
In this appendix, screenshots of collaborative problem solving tasks, not described in detail in this chapter, are presented. The tasks are Hexagons, Hot Chocolate, Plant Growth, Small Pyramids, Shared Garden, Sunflower, Warehouse, Light box.



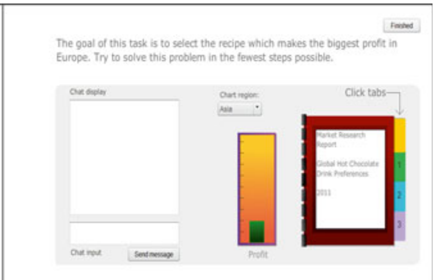
Student A View: Hexagons Page 5



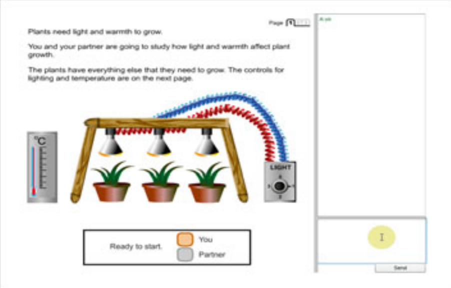
Student B View: Hexagons Page 5



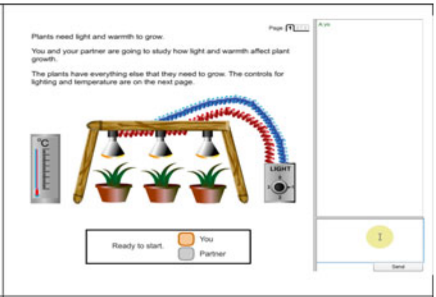
Student A View: Hot Chocolate Page 1



Student B View: Hot Chocolate Page 1



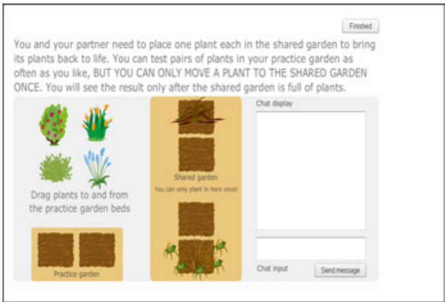
Student A View: Plant Growth Page 1



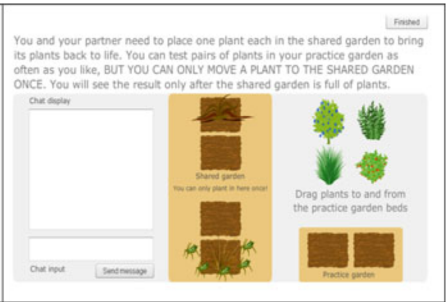
Student B View: Plant Growth Page 1

Student A View: Small Pyramids Page 1

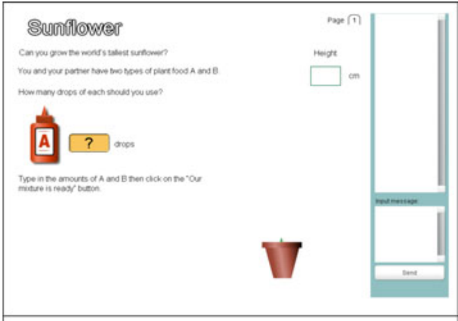
Student B View: Small Pyramids Page 1



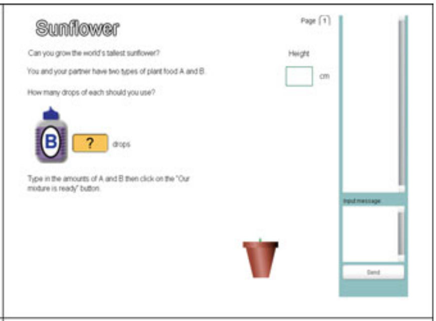
Student A View: Shared Garden Page 1



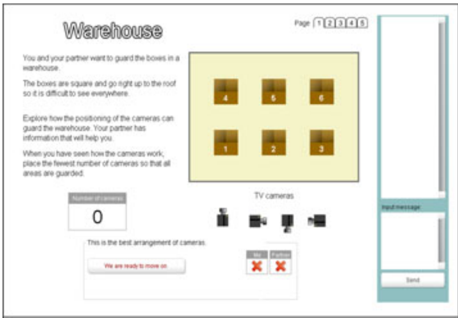
Student B View: Shared Garden Page 1



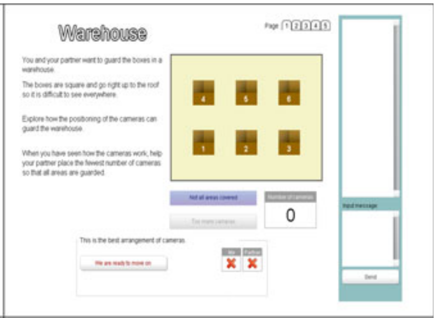
Student A View: Sunflower Page 1



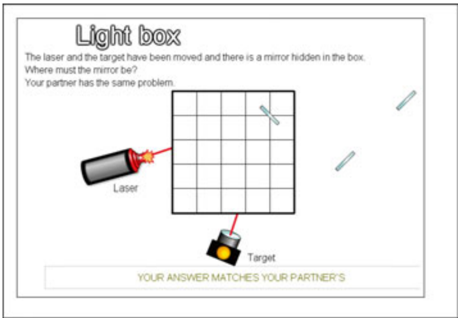
Student B View: Sunflower Page 1



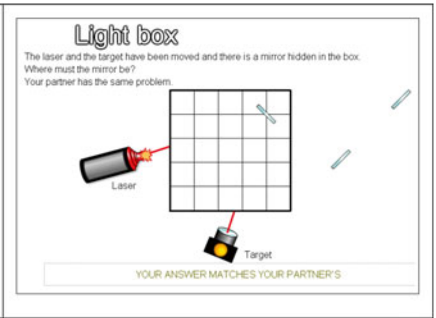
Student A View: Warehouse Page 1



Student B View: Warehouse Page 1



Student A View: Light Box Page 2



Student B View: Light Box Page 2

References

Adams, R., Vista, A., Scoular, C., Awwal, N., Griffin, P., & Care, E. (2015). Automatic coding procedures for collaborative problem solving. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills: Methods and approach* (pp. 115–132). Dordrecht: Springer.

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

Funke, J. (1991). Solving complex problems: Exploration and control of complex systems. In R. J. Sternberg & P. A. Frensch (Eds.), *Complex problem solving: Principles and mechanisms* (pp. 185–222). Hillsdale: Erlbaum.

Garofalo, J., & Lester, F. (1985). Metacognition, cognitive monitoring, and mathematical performance. *Journal for Research in Mathematics Education, 16*(3), 163–176.

Griffin, P. (2014). Performance assessment of higher order thinking. *Journal of Applied Measurement, 15*(1), 1–16.

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* (pp. 37–56). Dordrecht: Springer.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs: Prentice-Hall.

Petković, M. (2009). *Famous puzzles of great mathematicians*. Providence: American Mathematical Society.

Polya, G. (1945). *How to solve it* (1st ed.). Princeton: Princeton University Press.

Polya, G. (1957). *How to solve it* (2nd ed.). Princeton: Princeton University Press.

Schoenfeld, A. H. (1985). *Mathematical problem solving*. New York: Academic.

Vollmeyer, R., Burns, B. D., & Holyoak, K. J. (2006). The impact of goal specificity on strategy use and the acquisition of problem structure. *Cognitive Science, 20*, 75–100.

Zoanetti, N. (2010). Interactive computer based assessment tasks: How problem solving process data can inform instruction. *Australasian Journal of Educational Technology, 26*(5), 585–606.

Chapter 5
Platforms for Delivery of Collaborative Tasks

Nafisa Awwal, Patrick Griffin, and Sam Scalise

Abstract Prior to the inception of the ATC21S™ project, no single platform existed for the implementation of human-to-human internet interactive tasks. (The acronym ATC21S™ has been globally trademarked. For purposes of simplicity the acronym is presented throughout the chapter as ATC21S.) Given the global goals of the project, the assessment tasks and the professional development materials for administration and teaching interventions were constructed on several platforms. The characteristics and capacities, and the advantages and disadvantages, of these platforms are described in this chapter. The mathematics-based collaborative tasks were developed on a Hong Kong-based platform and migrated to the online testing framework of the Assessment Research Centre (ARCOTS) at the University of Melbourne. The reasoning-based tasks were developed directly on this platform. Another platform (FADS) was used for the development of the Learning in Digital Networks or ICT (LDN-ICT) tasks. An attempt was made to modify a serious game product but the platform requirements of the developer (Pixelearning) led to this attempt being abandoned. Issues such as automating assignment of login codes, report generation, and real-time synchrony are discussed in detail in this chapter. The chapter highlights factors that discriminate between the maintenance and management of a local system of assessment and teaching materials and the issues that need to be considered when maintaining and managing such a system on a global scale.

Overview and Design Rationale

The advancement of technology, the growing use of the Internet and the accessibility of the World Wide Web have had a major impact on education. The use of computers and other electronic devices now plays an important role in designing, delivering and

N. Awwal (✉) • P. Griffin
Assessment Research Centre, Melbourne Graduate School of Education,
University of Melbourne, Parkville, VIC, Australia
e-mail: n.awwal@unimelb.edu.au

S. Scalise
University of California, Berkeley, CA, USA

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_5

authoring educational assessments. The uses of web-based and closed computer-based platforms for delivering assessments has become common practice. Many of these systems are still limited in supporting delivery of traditional assessment methods such as multiple-choice questions, ‘fill in the blank’, and so on. The advantages of using real-time delivery media include remote accessibility (web-based), automated scoring, instant feedback, convenient data storage and so forth. Although these systems have accelerated the process of marking and analysing students’ responses, they have not yet provided any psychometric advantages over traditional paper-and-pencil-based assessments because they have not been used to capture data that link assessment to the students’ cognitive processes or decision making behaviour.

Although still in their relative infancy, however, computer-based assessments can be used to capture information-rich student performances. Unlike traditional test items, computer-based assessments can provide the means for capturing, recording, storing, processing and scoring data that reveal the processes by which students reach their answers. The process data derived through such assessments are considered richer than traditional data as they can describe the type, order and quantity of interactions with a task (Bennett et al. 2003; Greiff et al. 2012). This form of data is essentially collected through the capture of discrete mouse and keyboard events in a process discussed in more detail by Adams et al. (2015; Chap. 6). The technological requirements of such assessments are those for the capture, storage and processing of time-stamped click events. Various forms of client and server configurations can be used for this purpose. The key technologies used to develop the collaborative problem solving tasks in ATC21S are described in the following section. The technologies for the delivery of LDN-ICT tasks are described in the subsequent section.

Delivery of CPS Tasks

The current suite of collaborative problem solving (CPS) tasks for ATC21S is based on an established set of computer-based problem solving assessment tasks developed by Zoanetti (2010) for single-user administration. Their conversion into tasks for two users with differentiated screen views and real-time collaboration tools facilitated the testing of their capacity to elicit collaborative problem solving skills (refer to Care et al. 2015; Chap. 4 for more details).

The CPS assessment tasks developed were delivered through the online assessment system ARCOTS (Assessment Research Centre Online Testing System) of the Assessment Research Centre, University of Melbourne. The authentication process for students controlled access to the tasks. Since the delivery was online, students were able to access the tasks at their convenience via any client browser with a relatively up-to-date Flash Player plug-in.

The graphical components of the reasoning-based CPS tasks were designed using Adobe CS5 and programmed with ActionScript 3. The use of Flash limited the range of devices, such as tablets with iOS, that could be used for the assessment tasks. SmartFoxServer 2X[1] was chosen as the socket server technology enabling clients to share communication and object-manipulation data. SmartFoxServer 2X was chosen over other alternatives because of its availability in a free community version designed to support developers. Thus, all packages for libraries were incorporated into the development environment for use throughout the development of these tasks. To support the design of some of the animations in the tasks, motion tweening utilities were also used. In accordance with the requirements of the ATC21S project, the tasks were initially delivered in four different languages: English, Spanish, Finnish, and Dutch; the XML Strings Panel provided in the Flash IDE was used to embed the appropriate language, as determined by the client operating system, in each task. The tasks automatically detect the language settings of the user's web browser and IP location, and if the language is supported by the system, then all of the assessment tasks are shown in that respective language; else everything defaults to a common language which had been set to as 'English'.

The server components included the LAMP stack (Linux, Apache HTTP Server, MySQL and Perl/PHP/Python) as part of the multi-user architecture, implemented on a remotely hosted server. SmartFoxServer 2X was mounted to provide an open connection between the clients via the server. Use of such a socket application became essential because although single-user architecture might be able to support turn-based games, it would constrain design of tasks which rely on real-time (or minimal lag) updates for both users following activity by one or the other. A number of custom Java packages were developed and incorporated into the platform to supply some of the assessment task logic and to handle the flow of both-way data communication between clients and the MySQL database. The database – designed as a relational structure – and the application packages were configured to support the various target languages and output the resultant task view at the client's end.

Each of the curriculum-based CPS task was implemented as an Adobe SWF Flash 10 object embedded in a customised PHP page and delivered from the same Linux based web hosting server. The swf in the client's browsers was connected to a Flash Media Server for communications and synchronisation of shared objects between the collaborating users. The web server also provided an AMF gateway (which is a PHP based, open source package for handling server-side calls) to allow clients to log into collaborative session events. The task elements written in ActionScript 3 require remote shared objects on Flash Media Interactive Server[2] (FMIS). FMIS is necessary to support the communications and synchronisation of shared objects between the collaborating users and is considered a standard choice for streaming media and shared object synchronisation for swf-based clients. The language detection in this case relies on the language information sent by the browser to the server in the request header.

[1] SmartFoxServer is a middleware for creating multi-user applications, communities and games with Adobe Flash/Flex/Air, HTML5, Android, Unity3D, Apple iOS, Java etc. 2X is its community edition.

[2] Flash Media Interactive Server is from the Adobe Media Server family and supports multi-user media and games delivery on various devices with a wide range of deployment options.

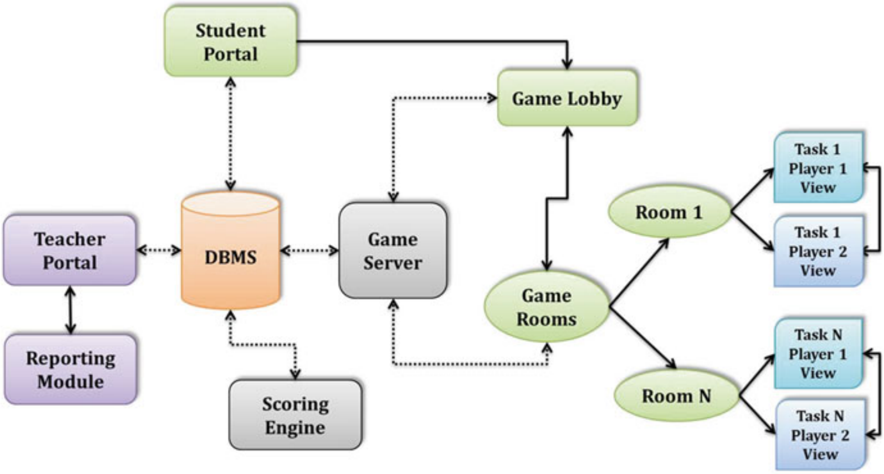


Fig. 5.1 Schematic diagram of the platform

The initial system was eventually expanded to incorporate Flash assessment tasks designed by third parties who used the FMIS as the socket server technology. Hence, an instance of FMIS was also connected on the Linux operating system as a service. The system was built to integrate with the established ARCOTS platform so that the authentication, scoring and reporting modules were common across all assessment environments. While it is not yet IMS QTI[3] standards compliant, and therefore not optimally interoperable, some modularity was built into the platform to help ensure that other forms of assessment tasks could be accommodated. Appropriate fonts for each of the target languages were also embedded in the chat messaging interface.

The assessment tasks were created to be executed in a similar fashion to that of a game within an online multi-user gaming architecture (Fig. 5.1). The collaborative environment followed some basic steps to give users access to the tasks. As mentioned earlier, access to tasks is controlled by an authentication process and limited to participating countries. Hence, students can log into the system using valid credentials (unique student identifier and a shared team code to pair collaborators in a session). On successful login, a student enters a virtual room, referred to as '*Lobby*' in the gaming environment and is presented with a set of CPS tasks for selection. Once students select a task to attempt, they are further provided with options to select a role for different views of the task (unknown to them) as presented for interaction. As soon as the students click on a user icon (avatar) to select their role, a virtual room is created dynamically on the basis of predefined rules (a combination of unique task identifier, shared key and other variations). Upon

[3] The IMS Question and Test Interoperability (QTI) specification defines a standard format for the representation of assessment content and results, supporting the exchange of this material between authoring and delivery systems, repositories and other learning management systems. It allows assessment materials to be authored and delivered on multiple systems interchangeably. It is, therefore, designed to facilitate interoperability between systems.

admission, the student is informed that the room is waiting for their pre-assigned partner to join them. If the student is the second of an assigned pair to click on a user icon, that student will enter the virtual room created by the first student, and the task contents will automatically be presented to both students with differing views and resources. While paired under a session, students can collaborate or operate among themselves and within the problem space. At the completion of a task, students are redirected back to the task-list dashboard or onto a self- and/or peer-assessment.

The whole process steps users through a shared session while collecting their responses for collaborative behaviour. The database logs detail students' responses including actions by students that trigger a change to the task on screen (buttons, text inputs etc.) and also events that don't trigger any change (for example, clicking on buttons which are not enabled). In short, any clicks or activities on screen by students are captured regardless of their effectiveness, because invalid, ineffective and tentative actions may prove to be more informative in later analysis. Classes were defined (PHP and Java were used) to record such responses and to keep track of how users are interacting with the system.

The system is also designed to deal with unexpected interruptions due to internet connection failures or other technical issues: in such events it allows students to continue access from their previous responses. During a collaborative session, if one student of the pair encounters a technical problem and closes their window (or hits refresh/back button on their browser), the system will allow both users to re-enter the task at the initial state of the last common page they shared. On return to the same task, the first returning user sees the role selection page with their previous role already selected and both users are automatically returned to the page on which they were last collaborating. If, at any point, one partner's session stops, the other partner should not be able to proceed to a new page: this can happen only when both partners confirm the next page request within a task. In addition, the other partner is provided with a system notification of partner loss. If a partner closes or refreshes their browser window, the system will detect this activity and – as long as the network connection remains active – inform the other partner of it. The tasks being collaborative, no user is allowed to move forward without corresponding partner.

Given the intended proliferation of tasks in the ATC21S project, programming re-usable client-side and server-side classes and/or packages with ActionScript 3 and Java became an important objective because it allows for efficient up-scaling. These classes managed the majority of processes common across the set of tasks available, including student login verification, real-time chat messages, data storage to the database, real-time sharing of information about task objects between students, and other aspects of game logic.

Delivery of LDN-ICT Tasks

A set of LDN-ICT tasks – different in theoretical context to the CPS tasks – was both developed and delivered through the Formative Assessment Delivery System (FADS) of the Bear Center, University of California, Berkeley. Like the CPS tasks, these were delivered online, but through FADS, using appropriate authentication protocol for participating countries in the ATC21S project. FADS had adhered to a model-based assessment system with reasonable Application Program Interfaces (APIs) and has some compliance with SCORM[4] and IMS QTI. SCORM is a collection of standards and specifications for web-based e-learning. It defines communications between client-side content and a host system called the run-time environment, which is commonly supported by a learning management system. SCORM also defines how content may be packaged into a transferable ZIP file called a "Package Interchange Format".

The LDN-ICT tasks were designed and developed with Adobe Flash in ActionScript 3. All of these Flash tasks created within FADS communicate with an intermediate Flash object, "integrator", which forwards the required requests to the backend database architecture. The integrator is defined by a flash parameter with the name "integratorUrl". The value contains the URL to the swf file that implements the integrator (e.g. http://berkeley.edu/somedirectory/integrator.swf).

All the LDN-ICT tasks were deployed as Adobe Flash swf object and contained a configuration XML that is used for rendering and scoring purposes. This XML needs to be stored by the backend and is delivered to the Flash task through the integrator. In the case of the FADS integrator, this communication is encrypted to avoid disclosure of correct and incorrect answers, but this is not an absolute requirement.

The set of LDN-ICT tasks was delivered only in English and Spanish, with slightly different content design for the respective countries using those languages. The language of the content was not reliant on browser language detection but merely on the task selection that is embedded with users' preferred language. The chat messaging interface was absent from these form of assessment tasks but FADS has used other forms of standard web applications, such as Google Docs, Webspiration Classroom, Kodu GameLab and other similar programs, that facilitate communication and sharing among collaborating groups. In addition, the tasks necessitate the use of external sites and are embedded with required resources. Unlike the CPS tasks, these assessments were designed to be played among two to four collaborating partners.

Most of these tasks are designed to be revisited by the student so that the student can review and alter their previous responses. When each task loads, it assembles an XML request to retrieve any prior responses and uses the integrator's defined function (e.g. getMyResponses) to retrieve past responses and submit the request to the hosting environment. When the responses are received, the integrator issues an event such as 'responsesReceived', to which the task responds by calling the integrator's function (e.g. getMyPreviousResponsesXML), which delivers the response XML. Some of these tasks are designed to incorporate a student's responses to other tasks. In those cases, the task calls another integrator function (e.g. getPreviousResponse) to retrieve relevant responses for the student. This function also handles unexpected interruptions to collaborative sessions caused by internet connection failures or other technical issues.

[4] Sharable Content Object Reference Model (SCORM) is a specification of the Advanced Distributed Learning (ADL) Initiative, which comes out of the Office of the United States Secretary of Defense.

A number of these LDN-ICT tasks rely on student-specific information stored in the hosting environment, such as the student's age, team assignment, or login credentials, to alter the content that is displayed and the availability of access to external websites. Tasks assemble a request for such information as a piece of XML and use the integrator's function (e.g. getDemographics) to retrieve the student's profile and relevant demographic information and submit that request to the hosting environment. When the information is received, the integrator issues an event like 'demographicsReceived', to which the task responds by calling the integrator's function, such as getMyDemographicXML, which delivers a similar XML fragment containing the requested information.

To gain access to the LDN-ICT tasks in a collaborative environment, the assessment system follows a similar multi-user architecture to that of the CPS tasks. Again, as in the case of the CPS tasks, access is restricted through authentication protocol. Collaborating students within the same team are presented with parallel views and resources in their respective browsers, but are allowed to progress into the task space at their own pace, without restrictions on collaboration with prospective partners. Students are allowed to complete the task in their own time and are not required to wait for collaborating partners. Students are made aware of their partners only through the use of the collaborative spaces they share. The tasks were designed to capture student responses or actions, both shared and unshared, within the task environment and external resources. The responses and/or actions mainly consist of various forms of user inputs (textual, graphical, multimedia, etc.) and retrieval of information through the range of resources provided both internally and externally. Any such activity within the task space is thus captured and fed into the backend of the FADS database in the appropriate format as defined by the content delivery mechanism. Again, due to the use of Flash, the range of useable devices was limited (e.g. PCs but not Apple iPads), and the reliance of these tasks on external resources and applications imposed further restrictions on the availability of those resources.

Lessons for an Integrated Collaborative Assessment Platform

This section will focus on some of the issues identified with the earlier assessment platforms and provide some suggestions for design choices on technologies as guidance for future implementation. As discussed in earlier sections, multiple platforms co-existed for delivering the range of assessments of the ATC21ST project. Having multiple platforms with differing technological requirements made it difficult to deliver examples of the ATC21S assessments (both CPS and LDN-ICT tasks) to a wide range of students. It was realised that existing know-how needs to be integrated with emerging technologies to introduce and consolidate multiple portals, automated scoring algorithm of student responses and the feedback mechanism to teachers; many of these were missed in the earlier version of the delivery platforms. Key considerations for such an integrated system should include access for large numbers of simultaneous users, the dispersed locations of collaborating users and devices in use, the multi-lingual capability content management system and results of user attempts provide feedback in real-time. The aim of such a feedback mechanism, similar to the one developed as part of the Assessment and Learning Partnerships project (Griffin 2000), is to allow teachers to monitor student progress over time and link that progress to successful teaching strategies.

Cloud based technologies for assessments are relatively new, but other e-business' (e.g. EBay, Amazon etc.) have commercialised their development efforts into services. This can be used as an alternative medium for interactive task delivery to recuperate some of the issues that earlier Flash tasks or technology may have imposed. Open standards, specifically the W3C standards, in such technologies can be adhered easily, making it a cost effective solution for such deployment. In addition, such technologies offer a variety of patterns and practices that can allow the design of any system to ensure it can be scaled to support a large number of users. Such technology can generate 10–40 times more transaction data compared to traditional online test items due to the nature of the interaction from these collaborative assessment tasks.

As discussed, a key criterion for this project was to develop interactive assessment tasks that use synchronous communications to provide support for a collaborative process. HTML5 has been gaining momentum over Flash in creating such applications. Unlike Flash, HTML5 is supported by all devices (such as Apple iPads) and is not constrained by licensing limits for the use of additional server based software. It can accommodate synchronous communication among thousands of users and thus allow for scalability on the level required by most projects. This option is likely to make financial support more viable for deployment in schools later.

A system based on these latest technologies will provide a consistent experience for users across all browsers and platforms. A disadvantage of these design choices is that HTML5 standards are not adequately supported by current browser versions that are more than two years old. However, while some schools may still be using older version browsers, the freely accessible latest versions do help to alleviate this problem and it is expected that all schools will upgrade to current HTML5 browsers over the coming years.

Implications and Future

Development of complex interactive assessment tasks poses logistical and pedagogical challenges, not only for the developers but also for potential users of the technology. Schools – the target locations of such assessment use – are usually restricted in their access to state of the art technology, including software, hardware, internet access and bandwidth. This gap between the available technology and student access to it is more profound in rural areas and other locations that are not technology-rich. As the need for teaching 21st century skills is recognised, there will be an increasing demand for such assessments, and schools are likely to become better equipped to meet their students' needs for the future.

To design challenging and rational tasks that requires concrete collaboration while preventing all single-player solutions is not easy. However, as new technologies and standards evolve, it is reasonable to expect that new collaborative tasks that make use of the functions described here will be conceived and created. Future plan need to include development of a range of new collaborative tools that can accommodate more features for collecting salient information on students' collaborative problem solving processes. As new schools and new countries take up the ATC21S system, there may be a need for the integration of more functionality and some scenarios may entail new development. To date, the development has been a learning process for the teams involved, and many ideas and concepts have been continually refined. With research needs and opportunities rapidly evolving, an even greater need to explore and extend the system is yet to unfold.

References

Adams, R., Vista, A., Scoular, C., Awwal, N., Griffin, P., & Care, E. (2015). Automatic coding procedures for collaborative problem solving. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills: Methods and approach* (pp. 115–132). Dordrecht: Springer.

Bennett, R. E., Jenkins, F., Persky, H., & Weiss, A. (2003). Assessing complex problem solving performances. *Assessment in Education, 10*(3), 347–359.

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

Greiff, S., Wüstenberg, S., & Funke, J. (2012). Dynamic problem solving: A new assessment perspective. *Applied Psychological Measurement, 36*(3), 189–213.

Griffin, P. (2000). *The profiles in practice: School reporting software*. Portsmouth: Heinemann.

Zoanetti, N. (2010). Interactive computer based assessment tasks: How problem-solving process data can inform instruction. *Australasian Journal of Educational Technology 2010, 26*(5), 585–606.

Chapter 6
Automatic Coding Procedures for Collaborative Problem Solving

Raymond Adams, Alvin Vista, Claire Scoular, Nafisa Awwal, Patrick Griffin, and Esther Care

Abstract This chapter examines the procedure followed in defining a scoring process to enable the reporting of individual student results for teachers to use in the classroom. The procedure begins with the identification of task features that match elements of the skills frameworks, and is followed by the generation of simple rules to collect data points to represent these elements. The data points are extracted from log files generated by students engaged in the assessment tasks and consist of the documentation of each event, chat and action from each student. The chapter includes examples of the process for defining and generating global and local (task specific) indicators, and examples of how the indicators are coded, scored and interpreted.

The development of coding and scoring of data generated when students engage in collaborative problem solving tasks is described. The data generated are captured in a process stream data file. Patterns of these data are coded as indicators of elements defined in the conceptual framework outlined in Hesse et al. (2015; Chap. 2) and the relative complexity of indicators is used in a scoring process. The scored data are then used to calibrate the tasks. The calibrations form the basis of interpretation and these are used in forming reports for students and teachers. Figure 6.1 summarises the entire process from task development to the reporting of student ability based on a developmental framework.

R. Adams (✉) • A. Vista • C. Scoular • N. Awwal • P. Griffin • E. Care
Assessment Research Centre, Melbourne Graduate School of Education, University of Melbourne, Parkville, VIC, Australia
e-mail: r.adams@unimelb.edu.au

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_6

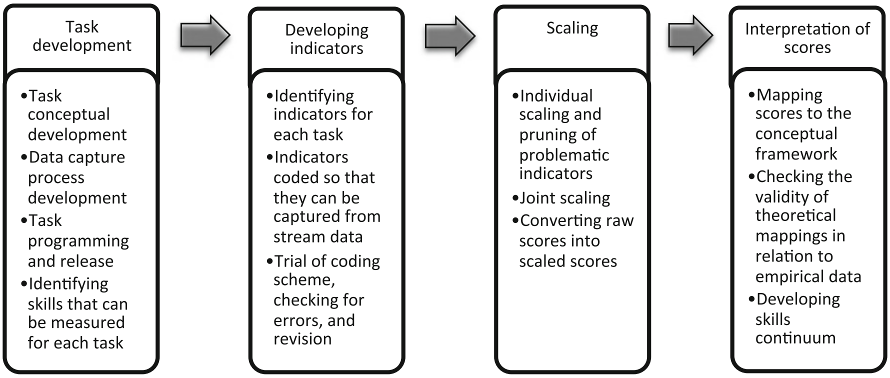


Fig. 6.1 Process overview from task development to interpretation of scores

Existing Approaches to Autoscoring

There is currently very little research regarding a scoring approach for collaborative problem solving. Our project has therefore focused on and adapted existing scoring processes for problem solving. The literature suggests that current processes mainly use a dichotomous success-failure scoring system which records whether the problem has been solved and ignores the cognitive procedures involved (Greiff et al. 2012). This type of system is simple to implement and works well for tasks which are designed to tap into specific problem solving skills. For example, a task where deductive reasoning is imperative for success can be scored dichotomously. An example of this style of dichotomous scoring can be observed in a project by Greiff et al. (2012) who have determined three measures which represent dynamic problem solving (DPS): Model Building, Forecasting and Information Retrieval. Each of these measures is scored across 11 DPS tasks and students are awarded a false (0) or true (1) score determined by their success or failure on the task. In contrast, the focus in the ATC21S™ project[1] is not only to determine whether students are succeeding at solving the tasks but to draw inferences about *how* students solve problems. While the assumption in traditional test design is that the attainment of the solution is the sole criterion, here the focus is on the process and quality of problem solving. A distinction needs to be made between what might be called simple problem solving tasks, using a dichotomous scoring process, dynamic problem solving, using a series of dichotomous scores, and complex problem solving, using rubrics and partial credit approaches.

The procedural aspects of problem solving (PS) have been considered important for some time (Polya 1945, 1957; Schoenfeld 1985). The framework proposed in the ATC21S project outlines five broad components that represent collaborative

[1] The acronym ATC21S™ has been globally trademarked. For purposes of simplicity the acronym is presented throughout this chapter as ATC21S.

problem solving (CPS) within social skills (participation, perspective taking, social regulation) and cognitive skills (task regulation and knowledge building). Within these five components, students are assessed on three tiered levels of ability across 19 specific elements. A set of assessment tasks, each tapping into different and overlapping skills, was developed in order to provide teachers with sufficient information to interpret students' capacity in CPS subskills, so that a profile of each student's performance can be developed for formative instructional purposes. In order to provide this interpretive feedback, there was a need to develop a robust automated scoring system which highlighted the procedural and developmental thinking processes that take place in CPS.

Design of Process Data Stream – Capturing and Identifying the Data

Many recent computer based PS tasks have been able to assess and record detailed interactions between the problem solver and the task environment, and thereby capture salient solution processes in an unobtrusive way (Zoanetti 2010; Bennett et al. 2003). Their recorded input can be linked to the cognitive skill level and development of students and used to evaluate the process and efficiency with which problem solvers complete tasks (Pelligrino et al. 2001; Williamson et al. 2006). Within the current CPS framework, actions and chat, and the placement of these, can be scored.

In order to record descriptive, purposeful actions, the CPS assessment tasks were designed to capture detailed interactions between problem solvers working as a dyad as well as between the individual problem solver and the task. In the context of computer based assessments, the files generated for the automatic records of these types of student–task interactions are referred to as a 'session log file.' They contain free-form data referred to as 'process stream data.' The log files were stored as free-form text files with delimited strings of text or in database architecture. In this instance, MySQL database architecture was used for recording the interactions with the task environment thereby describing relevant solution processes in an unobtrusive way (Bennett et al. 2003).

In the context of these assessment tasks, process stream data describe distinct key strokes and mouse events such as typing, clicking, dragging, cursor movements, hovering time, action sequences and so on. In the database, each discrete action is recorded with a corresponding timestamp. A timestamp refers to the time at which an event was recorded by the system into a log file and reflects, or is close to, the time of the event itself. To ensure that the data captured can be presented in a consistent format, allowing relatively easy comparison of two different records and tracking progress over time, a sequential numbering of events is used in a consistent manner. In this way, timestamps enable detailed analysis of action sequences and inactivity. This ensures further transparency for data storage and the sequential logging of events for data processing. These forms of time-stamped data have been

Table 6.1 Examples of common events defined from the process stream data

Event type	Process stream data format	Explanation of data captured
Session start	Student ***student_id*** has commenced task ***task_id***	Records the start of a task with student and task unique identification
Session finish	Student ***student_id*** has completed task ***task_id***	Records the end of a task with student and task unique identification
Chat text	Message: "*free form of message using the chat box*"	Captures the contents of the chat message the students used to communicate with their partner
Ready To progress	Requested to move to page: ***page_id***	Indicates whether the student is ready to progress or not, and records the navigation endpoint which they are ready to progress to for multipage tasks
Other click	Screen x coords: ***x_coordinate***; Screen y coords: ***y_coordinate***;	Captures the coordinates of the task screen if the student has clicked anywhere outside the domain of the problem

referred to variously as "log-file data" (Arroyo and Woolf 2005), "discrete action protocols" (Fu 2001), "click-stream data" (Chung et al. 2002) and "process data" (Zoanetti 2010). For our purpose, we use the term 'process stream.'

Each task has a variety of events that can occur, categorised into two types: common and unique events. As the name suggests, 'common' classifies the universal nature of the process stream events and applies to all collaborative assessment tasks. Examples of these events can been seen in Table 6.1. They include indications of the beginning and end of a task, system confirmation messages of user actions, navigational system messages for multiple page assessment tasks, free-form chat messages for communication with partners, or variations of these.

Unique events within the process stream data are not common across assessment tasks. They are unique to specific tasks due to the nature of the behaviours and interactions those tasks elicit. These data are defined using event types to match specific requirements that may arise only in a particular interactive problem space. Examples of such events for the Laughing Clowns task, illustrated in Fig. 6.2, are presented in Table 6.2 (for a detailed explanation of this task see Care et al. 2015; Chap. 4).

The accumulation of the different types of process and click stream data collectively forms the process data stream, accumulated and stored in files commonly referred to as *session logs*. An excerpt of a session log for the Laughing Clowns task can be seen in Fig. 6.3, which represents the events that occurred for one team (two students) while playing the task. Both common and unique event types of process stream data were captured in string format as shown in Tables 6.1 and 6.2. Process stream string data were recorded in the MySQL database as a single row and tagged with corresponding student identifier, task identifier, page identifier and role allocation of the acting student in the collaborative session with time-stamping and appropriate indexing.

To facilitate collaboration, a chat box tool was used (see Care et al. 2015) as the messaging interface for communication between respective partners during the collaborative sessions. This enabled the students to explore and learn about their respective resources and share or report information to each other. The chat box tool

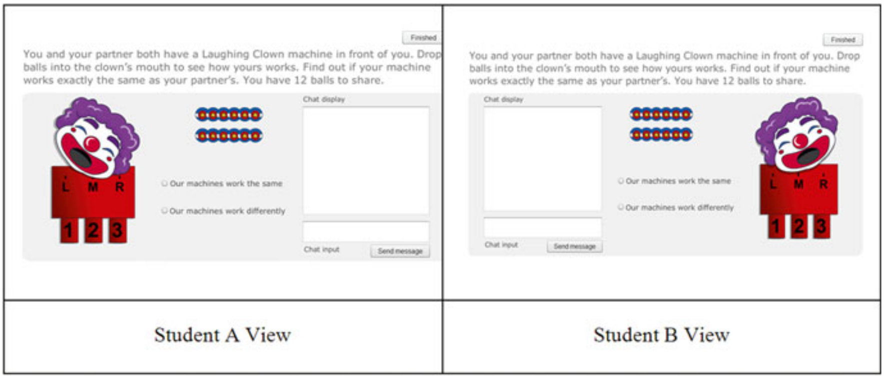


Fig. 6.2 Screenshots from the Laughing Clowns task

Table 6.2 Examples of unique events defined from the Laughing Clowns task within the process stream data

Event type	Process stream data format	Explanation of data captured
StartDrag	startDrag: *ball_id*; *x,y coordinates of the ball at the start of the drag*	Records the identifier of the ball which is being dragged by the student, and its coordinates
StopDrag	stopDrag: *ball_id*; *x,y coordinates of the ball at the end of the drag*	Records the identifier of the ball which is being dragged by the student and its coordinates at the end of the drag
DropShute	dropShute*PosofShuteId*: *ball_id*; *x,y coordinates of the ball when it was dropped*	Records the identifier of the ball, its coordinates and the value of the clown head shute when it was dropped by the student
Check box	Selection Value: *option_value*	Captures data if students agree or disagree on how their machines work

captures text exchanged between students and was captured in string data format. All chat messages generated by users and the system appeared in the tool and were recorded with a corresponding timestamp.

Defining the Indicators

Each task was examined for indicative behaviours of identifiable cognitive and social skills that can be captured algorithmically. These skills were identified through actions, chats or a combination within the process stream. The behaviours that were observed in the process stream data were used as indicators of

record	actor_pid	team_id	task_id	page	player_id	event	data	timestamp	Events
127988	student0951	sng0076	103	1	A	start	Task started is 103	26/09/11 16:28	Checkbox
127989	student0952	sng0076	103	1	B	start	Task started is 103	26/09/11 16:28	startdrag
127995	student0951	sng0076	103	1	A	action	startDrag:ball1:410:35	26/09/11 16:29	stopdrag
127996	student0951	sng0076	103	1	A	action	stopDrag:ball1:188:129	26/09/11 16:29	dropshute
127997	student0951	sng0076	103	1	A	action	dropShuteR:ball1:188:129	26/09/11 16:29	chat
128001	student0952	sng0076	103	1	B	chat	You put where ?	26/09/11 16:29	start/end
128015	student0951	sng0076	103	1	A	chat	i put on r	26/09/11 16:29	
128017	student0951	sng0076	103	1	A	chat	landed on 1	26/09/11 16:29	
128019	student0952	sng0076	103	1	B	action	startDrag:ball7:410:85	26/09/11 16:29	
128021	student0951	sng0076	103	1	A	action	startDrag:ball10:485:85	26/09/11 16:29	
128034	student0952	sng0076	103	1	B	action	stopDrag:ball7:792:133	26/09/11 16:29	
128035	student0952	sng0076	103	1	B	action	dropShuteM:ball7:792:133	26/09/11 16:29	
128038	student0951	sng0076	103	1	A	action	startDrag:ball9:460:85	26/09/11 16:29	
128039	student0951	sng0076	103	1	A	action	stopDrag:ball9:102:132	26/09/11 16:29	
128041	student0951	sng0076	103	1	A	action	dropShuteL:ball9:102:132	26/09/11 16:29	
128048	student0951	sng0076	103	1	A	chat	all of it land on 1	26/09/11 16:29	
128049	student0952	sng0076	103	1	B	chat	I never see where it landed	26/09/11 16:29	
128053	student0952	sng0076	103	1	B	action	startDrag:ball8:435:85	26/09/11 16:29	

Fig. 6.3 Excerpt from a process stream log file for the Laughing Clowns task

cognitive and social skills as defined in Hesse et al. (2015). These indicative behaviours were then coded into rule-based indicators that can be extracted from the task process streams through an automated algorithmic process similar to that described by Zoanetti (2010). Zoanetti showed how process data (e.g., counts of actions) could be interpreted as an indicator of a behavioural variable (e.g., error avoidance or learning from a mistake) (see Table 3 in Zoanetti 2010). For example, in the Laughing Clowns task, a count of the 'dropShute' actions (dropping the balls into the clown's mouth) can indicate how well the student managed their resources (the balls).

Coding

The coded indicators became the primary source of data for the scoring process. The indicators were classified into two main types: those that occur only in specific tasks and those that can be observed in all tasks. Indicators that can be captured in all tasks are labelled 'global'. They included total response time, response time to partner questions, action counts, and other behaviours that were observed regardless of the task. Indicators that were task-specific were labelled 'local'. There were two categories of local indicators: direct and inferred. Direct indicators represented those that can be identified clearly, such as a student performing a particular action. Inferred indicators related to such things as sequences of action/chat within the data. Patterns of indicators were used to infer the presence of behaviour which is indicative of elements in the conceptual framework (see Hesse et al. 2015; Chap. 2). Within these indicators there were differences in intensity or patterns that provided additional information about the relative complexity of the indicated behaviour.

Each indicator was coded with a unique ID code. Using the example of the unique ID code 'U2L004A', 'U2' represents the Laughing Clowns task, 'L' indicates that it is a 'local' indicator specific to that task ('G' would represent that it was a global indicator that could be applied to all tasks), '004' is a numerical code specific to this indicator which is provided for ease of referencing and is sequential within each task (in this case 004 it was the fourth indicator created for the task) and 'A' indicates that this indicator is applicable to student A.

To capture the required data, once the indicators are identified they need to be defined in the process stream through programming algorithms. Each of the scoring algorithms takes process stream data (produced by the events of the participants in different tasks) as input and produces relevant output defined by the rule for the corresponding indicator. For example, if capturing the quantity of interaction within a task, the algorithm would count the occurrences of the event 'chat' in the process stream. The output for this indicator would be the numerical value representing the frequency of the chat. Table 6.3 outlines some exemplar algorithms. The first column in the table represents the indicator name. Details of the scoring rule for each indicator are described in column two. The third and fourth columns elaborate the algorithm and its output respectively.

The outputs from each of the indicators based on the algorithms are saved in a 'coded file'. The coded file presents the output values relevant to the algorithm. For example, if the indicator observes a count of actions, the raw numerical value will be present in this file. Indicators highlighted in yellow in Fig. 6.4 are still in raw counts (or frequencies). These indicators are later converted into either a dichotomy or partial credit.

Mapping

Each indicator was mapped onto an element of the conceptual framework (outlined in Hesse et al. 2015; Chap. 2). which consists of five strands – three comprising the social aspect and two comprising the cognitive aspect. The main purpose of this mapping process was to identify an underlying skill. To reduce judgment error in the mapping process, it was undertaken several times by different teams. An iterative process was used. Several panels of researchers reviewed the indicators and mapped them onto the conceptual framework. The process was repeated for each set of indicators within each task until a stable allocation was agreed upon. When the changes and revisions to the allocation of indicators to elements fell to a minimum, the element mapping was then considered to be stable and the interpretation process proceeded to the next step. As an example, the indicator U2L004A records whether a student covers all positions with their balls. This is assessed by the presence of three 'dropShute' actions in the process stream for student A – one for each of the three positions L (left), M (middle), and R (right). This indicator was mapped onto systematicity in the framework, suggesting that the student had explored the task through a strategic sequence of actions. An excerpt from a session log on how this is captured can be seen in Fig. 6.5.

Table 6.3 Example of algorithms to the corresponding indicator

Indicator name	Details	Algorithm	Output
U2L004A	Systematic approach. All positions have been covered.	Step 1: Find all drop ball occurrences captured as dropShute and their corresponding positions as dropShuteL, dropShuteR, dropShuteM.	Count values
U2L004B	Scoring rule: threshold value.	Step 2: Then count all the occurrences of the action recorded under 'dropShute' and their unique positions from the log.	
	Task name: Laughing Clowns.	Step 3: Increase the value of the indicator by one if one or more 'dropShute' occurs in the form of dropShuteR, dropShuteL, or dropShuteM.	
		Step 4: If the total number of unique dropShutes (dropShuteR, dropShuteL, and dropShuteM) from the log is less than three then the value of the indicator is defined as −1 to indicate missing data.	
Global001A	Acceptable time to first action given reading load.	Step 1: Find the starting time when a student joins a collaborative session.	Time
Global001B	Time (in seconds) spent on the task before first action (interpreted as reading time)	Step 2: Find the previous record of the first action.	
	Scoring rule: Threshold time.	Step 3: Find the time of that previous record (from step 2).	
		Step 4: Calculate the time difference obtained (from step 1 and step 3), indicating the time before first action.	
Global005A	Interactive chat blocks: Count the number of chat blocks (A, B) with no intervening actions. Consecutive chats from the same player counts as one (e.g., A,B,A,B = 2 chat blocks; A,B,A,B,A,B = 3 chat blocks; AA,B,A,BB = 2 chat blocks)	Step 1: Find all the consecutive chat from student A and B without any intervening action from A or B. Treat two or more consecutive chats from a single student as one chat.	Count values
Global005B	Scoring rule: threshold number.	Step 2: Increase the value of the indicator by one if one block is found.	

teamID	studentID	W8L101A	W8L102A	W8L110A	W8L313A	W8L314A	W8L315A	W8L320A	W8L213A	W8L211A
aus1141	student1281	1	0	0	1	1	0	4	1	1
aus1151	student1301	0	0	0	1	0	1	3	1	0
aus1152	student1303	1	0	0	1	1	0	31	1	0
aus1162	student1323	0	0	0	1	0	0	1	1	0
aus1288	student1575	1	0	0	1	1	0	13	1	1
aus1292	student1583	1	0	0	1	1	1	0	1	1
aus5001	student5001	1	2	0	1	1	0	4	1	1
aus5008	student5015	1	3	1	1	1	0	2	1	1
aus5092	student5183	1	2	1	1	1	0	0	1	1
aus5104	student5207	0	0	0	1	0	1	0	1	1
aus5106	student5211	0	0	0	1	0	1	0	1	1
aus5183	student5365	1	5	1	1	1	1	7	1	1
aus5189	student5377	0	0	0	1	0	1	16	1	1
aus5191	student5381	0	0	0	1	1	1	12	1	1
sng3008	sngstud3015	1	0	0	1	1	0	4	1	1
sng3012	sngstud3023	0	0	0	1	0	0	2	1	1
sng3120	sngstud3239	1	0	0	1	1	0	8	1	1
sng3122	sngstud3243	1	7	1	1	0	1	7	1	1
sng3124	sngstud3247	1	0	0	1	0	1	1	1	1
sng3127	sngstud3253	1	6	0	1	1	1	0	1	1
sng3202	sngstud3403	1	0	1	1	0	0	6	1	1
sng3203	sngstud3405	1	0	0	1	1	0	2	1	1
sng3210	sngstud3419	1	2	0	1	1	0	11	1	1
sng3307	sngstud3613	1	4	1	1	1	0	22	1	1
sng3312	sngstud3623	1	8	1	1	1	1	9	1	1
sng3313	sngstud3625	1	1	0	1	0	1	9	1	1
sng3359	sngstud3717	1	2	1	1	1	1	10	1	1
sng3365	sngstud3729	1	0	1	1	1	1	15	1	1
sng3450	sngstud3899	1	0	0	1	1	0	13	1	1
sng3452	sngstud3903	1	0	0	1	0	0	3	1	0
sng3458	sngstud3915	1	1	0	1	1	1	24	1	1
sng3468	sngstud3935	1	3	1	1	1	0	9	1	1
sng3476	sngstud3951	0	0	0	1	0	0	18	1	1
sng3480	sngstud3959	1	0	0	1	0	0	3	1	1

Fig. 6.4 Excerpt from a coded data file

Filtered by student A

127988	student0951	sng0076	103	1 A	start	Task started is 103	26/09/11 16:28
127995	student0951	sng0076	103	1 A	action	startDrag:ball1:410:35	26/09/11 16:29
127996	student0951	sng0076	103	1 A	action	stopDrag:ball1:188:129	26/09/11 16:29
127997	student0951	sng0076	103	1 A	action	dropShuteR:ball1:188:129	26/09/11 16:29
128015	student0951	sng0076	103	1 A	chat	i put on r	26/09/11 16:29
128017	student0951	sng0076	103	1 A	chat	landed on 1	26/09/11 16:29
128021	student0951	sng0076	103	1 A	action	startDrag:ball10:485:85	26/09/11 16:29
128038	student0951	sng0076	103	1 A	action	startDrag:ball9:460:85	26/09/11 16:29
128039	student0951	sng0076	103	1 A	action	stopDrag:ball9:102:132	26/09/11 16:29
128041	student0951	sng0076	103	1 A	action	dropShuteL:ball9:102:132	26/09/11 16:29
128048	student0951	sng0076	103	1 A	chat	all of it land on 1	26/09/11 16:29
128059	student0951	sng0076	103	1 A	action	startDrag:ball11:510:85	26/09/11 16:29
128065	student0951	sng0076	103	1 A	action	stopDrag:ball11:147:144	26/09/11 16:29
128067	student0951	sng0076	103	1 A	action	dropShuteM:ball11:147:144	26/09/11 16:29
128076	student0951	sng0076	103	1 A	chat	from wherE?	26/09/11 16:29
128077	student0951	sng0076	103	1 A	chat	on m?	26/09/11 16:29
128079	student0951	sng0076	103	1 A	chat	from M?	26/09/11 16:29
128092	student0951	sng0076	103	1 A	action	radioSelection:1	26/09/11 16:30
128099	student0951	sng0076	103	1 A	action	radioSelection:1	26/09/11 16:30
128101	student0951	sng0076	103	1 A	action	radioSelection:1	26/09/11 16:30
128102	student0951	sng0076	103	1 A	action	radioSelection:1	26/09/11 16:30
128105	student0951	sng0076	103	1 A	action	radioSelection:0	26/09/11 16:30
128107	student0951	sng0076	103	1 A	action	startDrag:ball4:485:35	26/09/11 16:30
128113	student0951	sng0076	103	1 A	chat	then for R?	26/09/11 16:30
128116	student0951	sng0076	103	1 A	chat	and L?	26/09/11 16:30
128124	student0951	sng0076	103	1 A	chat	then different i think	26/09/11 16:30
128128	student0951	sng0076	103	1 A	action	stopDrag:ball6:535:35	26/09/11 16:30
128129	student0951	sng0076	103	1 A	action	startDrag:ball6:535:35	26/09/11 16:30
128130	student0951	sng0076	103	1 A	action	stopDrag:ball6:201:123	26/09/11 16:30
128131	student0951	sng0076	103	1 A	action	dropShuteR:ball6:201:123	26/09/11 16:30

Fig. 6.5 Excerpt from a process stream log file for the Laughing Clowns task

Scoring

Indicators can be thought of as the equivalent of items in a conventional test. In order to obtain an estimate of student ability from the scored indicators it is necessary that the status of one indicator does not affect or depend on the status of others. Requiring indicators to be stochastically independent also avoids the complexity of scoring the absence of an indicative behaviour when it is dependent on another event. For instance, if indicator 002 is dependent on indicator 001, and both are dichotomous, the assessment of indicator 002 = 0 will differ depending on whether indicator 001 = 0 or 1.

Dichotomously Scored Indicators

Most indicators of behaviours in the AC21S tasks are designed to be indicative only of the presence or absence of an observable behaviour. This would provide for each student a coded value of '1' to the indicator if it is present and a coded value of '0' to the indicator if it is absent. Through the forcing of most of the indicators into a dichotomy, the interpretation of indicators becomes simpler than is necessary for partial credit coding and scoring. In the Laughing Clowns task, for example, a player needs to leave a minimum number of balls for his/her partner in order for the task to be completed successfully. If the process data shows that this minimum number was satisfied, the indicator can be scored as 1. If it is not satisfied, it is scored as 0.

Frequency-Based Indicators – Partial Credit Scoring

In cases where a particular indicative behaviour is monitored for frequency of occurrence, recording the frequency counts is useful (as indicated in Table 6.3), especially when the cut-off for a qualitatively differentiable interpretation of the behaviour is not clear. For example, the total time taken on a task and the time taken for a player to respond to a partner query can range from a few seconds to several minutes. A dichotomy-based score cannot capture the subtlety of differences in such a case. In the Laughing Clowns example given above, the cut-off value is well-defined because success on this task is impossible beyond the minimum number of balls retained. However, in other tasks this situation may not be recordable in such clear-cut values. There will be an intuitive interpretation that more errors mean less problem solving ability, but it might not be clear where to place a cut-off point for scoring purposes. In these situations, the counts of indicative behaviour are recorded and used to construct a frequency distribution of values for later scoring.

Frequency-based indicators need to be converted into polytomous score values by setting cut-off or threshold values. The distribution typically takes the form similar

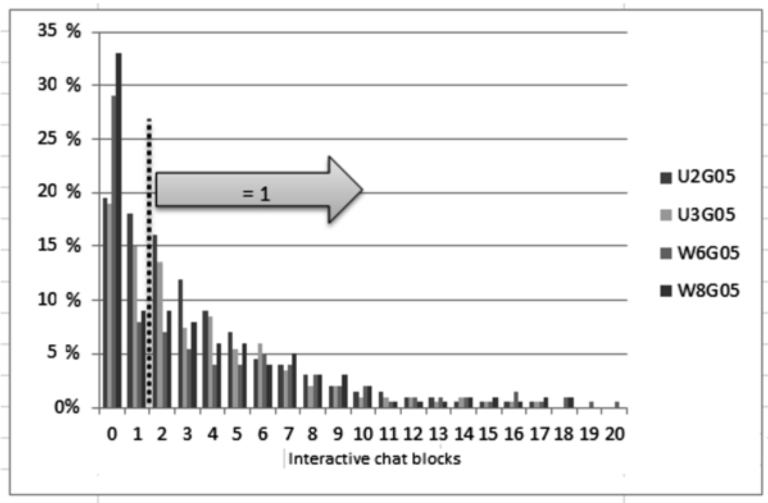


Fig. 6.6 Frequencies of inferred interactive chat blocks across four tasks for setting dichotomous categories

to an exponentially decreasing function or a unimodal function with a positive skew. An example of a decreasing function is shown in Fig. 6.6, where the distribution of inferred interactive chat blocks (chat A-chat B-chat A) for four tasks is illustrated. It shows a similar pattern of decreasing numbers of blocks, although the rate of decrease differs among the tasks to some degree. This type of distribution is scored by deciding where to put a cut-off point that divides the values into a dichotomy (high-low performance levels). If the cut-off value is set at 2, students who have interactive chat blocks of 0–1 get a score of 0, while those who have more chat blocks ($n \geq 2$) get a score of 1. The dichotomous scores can then be interpreted similarly to the presence-absence type of indicators where chat blocks ≥ 2 are taken as evidence of interaction (conversely, less than 2 chat blocks would be taken as insufficient evidence of interaction).

A second example, illustrated in Fig. 6.7, shows the distribution of response time on a question for the Hot Chocolate task. In this example, the mode is around 12 s, with the majority of elapsed time measures falling between 6 and 20 s. Deciding which range of values is qualitatively better is more difficult than in the previous example. Unlike the first example, where the scores were dichotomous, a unimodal distribution can have partial credit assigned to more than two different value ranges. Deciding the various value ranges and their score conversion equivalents can be done using empirical distributions and information obtained from relevant literature. For example, the period that elapses between chat and a following action could be regarded as ‘wait time’ and, although the concept of wait time in collaboration differs in intention and meaning to the ‘wait time’ in the literature, it can be used as a guide. The original concept of ‘wait time’ in a classroom setting refers to the time between a teacher-initiated question and a response from students (Rowe 1972). In

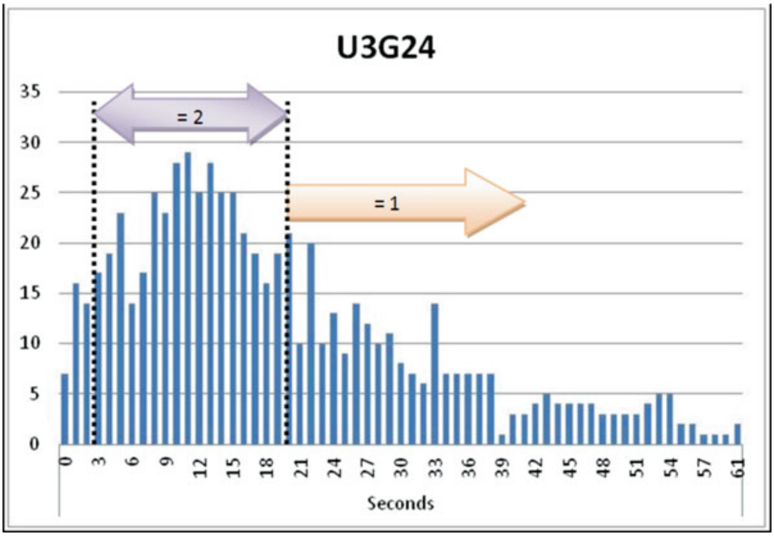


Fig. 6.7 Example histogram of a response time indicator for polytomous categories in indicator U3G24

that field of study, Tobin (1987) and Stahl (1990) suggested a minimum of three seconds wait as a threshold for positive student outcomes, such as increased chances of correct responses and higher cognitive levels of response. The context of their 'wait time' is different from the online setting, and their method of measurement was different from that of the collaborative tasks in ATC21S, but their concept provides a possible lower threshold for a reasonable score bracket (e.g., 0–3 s = 0, 3–20 s = 2, >20 s = 1).

Due to the unique nature of the ATC21S scoring approach, there was very little existing literature that could be used as a guide in setting the cut-off values for most of the process stream data. Since the empirical data for this variable were being captured for the first time in this project, setting the threshold cut-off values and assigning the partial credit scores was necessarily exploratory, and adjustments were made iteratively after calibration and interpretation. Setting the initial cut-off values was a precursor to calibration. The values were regarded as tentative descriptions of (qualitative) levels which were then checked for model fit and meaning during the calibration and scaling process.

Evidence of Collaboration Within Indicators

The evidence of collaboration in a task is primarily based on communication between the players. But it is more than simple communication. Student communication is not necessarily collaborative, or even cooperative. Such an interpretation at best would be simplistic and at worst incorrect. The ATC21S (Hesse et al. 2015; Chap. 2), and the PISA (OECD 2013) definitions of collaborative problem solving are clearly more nuanced than this. Using such simplistic definitions cannot help teachers develop their students' skills. Collaboration involves sharing, perspective taking, joint planning, decision making and shared goals. This cannot be summarised by a single indicator – 'students communicated'. It will involve both direct and indirect communication.

Indirect communication is inferred through actions that can be observed by collaborative partners. With this in mind, a specific approach to capturing chat was adopted. In the problem solving context, portions of the messages were recorded using a series of identifiable keywords. For collaboration, the *presence* of chat was recorded and the content of the chat was not taken into account. Chat linked to action – pre and post a chat event – was used to infer collaboration. This approach had the advantage of simplifying the data collection directly from the process stream while recognising the complexity of the collaboration itself. The presence/absence of chat, coupled with response time and action sequence data (i.e., when the chat occurred with respect to other actions or events), allowed a process to be used to infer collaboration. It was cross-checked by a separate panel of approximately 20 graduate students directly interpreting patterns of chat and action. This process made it clear that a simplistic APP approach which merely identifies the presence of communication is unlikely to enable collaboration to be accurately inferred.

There were several combinations of chat and action that could be interpreted as evidence of collaboration. Communication was inferred from patterns of chat or a combination of chat and action. If there was a presence of chat in the Laughing Clowns task after the last ball had been used and before the question had been answered then it was inferred that the students were discussing the potential answer. This was supported by the analyses of chat content.

The pattern of player-partner (A-B) interaction was also important to capture. For every pattern of chat-action possibility, player-partner combinations were also captured. The length (and hence the number of combinations) of player-partner interaction is unlimited (i.e., A, B, A, B, B, etc.). Hence, a limit of three sequences was adopted. With this limit in place, only the following player-partner combinations were possible: (1) A, B, A; (2) A, B, B; and (3) A, A, B. These combinations apply only to the action of the initiating student (A). Each student was coded separately in the data file, so the perspective changed when the other student (B) was scored. Only an interaction that was initiated by a student was scored for each student (i.e., we only scored for A the player-partner combinations that began with A, and vice-versa). Examples of combinations of interactions that can be captured are summarised in Table 6.4. In this table, the type of interaction (column 1) refers to all possible combinations of chat and action in a three-event block; the perspective (columns 4 and 5) refers to the sequence of player interaction (column 3) for these blocks from the perspective of the scored player (thus, it always begins with the scored player).

Table 6.4 Examples of inferred interactive chat-action combinations

Type	Measurement	Combination	Perspective from student A	Perspective from student B
Interactive **chat-action-chat** blocks	count	player + player + partner	AAB	BBA
	count	player + partner + partner	ABB	BAA
	count	player + partner + player	ABA	BAB
	count	player + player + player	AAA	BBB
Interactive **chat-action-action** blocks	count	player + player + partner	AAB	BBA
	count	player + partner + player	ABA	BAB
Interactive **chat-chat-action** blocks	count	player + partner + partner	ABB	BAA
	count	player + partner + player	ABA	BAB
Interactive **action-action-chat** blocks AAC	count	player + partner + partner	ABB	BAA
	count	player + partner + player	ABA	BAB

Defining the Skills Progression

After the rule-based indicative behaviours were identified, coded, and scored, the empirical data were examined to determine whether the mapping was consistent with the relevant skill in the conceptual framework (Hesse et al. 2015). This preliminary empirical analysis was undertaken to check if the relative difficulty of each indicator was consistent with the skill levels in the conceptual framework (Hesse et al. 2015). For example, an indicator that was interpreted and mapped to a simple level of participation in a task was expected to be less difficult (i.e., have a higher probability of being observed) than an indicator matched to systematic and exhaustive participation in optional activity in the problem space (a lower probability of being observed). Indicators were also reviewed by a panel to check the mapping of each indicator was relevant to the skill it was intended to measure. This panelling process also refined the definition of each indicator so that there is a clear link between the algorithm and the measurement construct. For example, an indicator algorithmically defined as "number (count) of resets (for the game)" can be refined and specified by extending the definition with "exploration activity and initial understanding of problem space". The refined conceptual descriptors were completed for all indicators independent of the empirical quantification of the item's relative position along the construct continuum (i.e., before they were placed into a hierarchical order of item difficulty [delta] based on a scaling under the Rasch

Model). After the indicators were ordered, based on empirical parameter estimates of their deltas, the hierarchy of the descriptors was again assessed to check that they make sense within a broader collaborative problem solving framework. This review process was completed several times to ensure that the conceptual descriptors are supported by empirical item location, which in turn informs the construct continuum. In the same process, the review clarifies which items have deltas that do not fit the theoretical model, and thus are not informative or meaningful within the overall structure of the construct.

After the skills progression was developed, levels of progression were identified in order to help teachers to cluster students more effectively and aid their instruction of CPS skills. The indicators were split into their two dimensions – social or cognitive – based on their previous mapping. Cognitive and social dimensions were each assessed independently to define a continuum and levels within each. Skills within each dimension were identified to represent the progression from novice to expert.

At this point, indicators which proved to have little value or influence on the interpretation were removed. The deletions were based on extensive item review, psychometric characteristics, and mapping to the theoretical continuum. The pruning is due to some indicators not matching the conceptual framework vis-a-vis their placement as expected from the theoretical progression. Also pruned were some indicators with coding issues which couldn't be resolved after extensive review.

Multiple calibrations allowed for comparison and analysis of item parameters. The stability of these parameters remained, even if the number of indicators was reduced considerably. As a result of the refinement process, the number of indicators was reduced from over 450 to fewer than 200. The removal of poorly 'fitting' indicators reduced the standard errors of the item parameters, while maintaining the reliability of the overall set.

Challenges and Future Directions

Even the most successful projects have lessons from which we can learn. The purpose of this section is to describe some of the lessons learned during development and deployment of the collaborative problem solving task design and delivery. What follows are descriptions of measures that are recommended as good practice to improve the design and implementation of such assessment tasks and data structure.

Design of the session log is crucial. The importance of leveraging complex and interactive assessment tasks not only to implement assessment delivery but also to establish automated scoring has been highlighted by many researchers (Mills et al. 2002; Williamson et al. 2006). The format in which data points are captured ensures efficient interpretation of user responses for establishing reliable scoring rules based on the evidence of interactive patterns from the logs. To validate the scoring rules, log files should be structured to allow human interpretation without obscuring their understanding. For example, each user action or response should be recorded as separate attributes in human readable format and as single instances with corresponding user identification, task and present state, timestamp, record index and other data as required for the task. In addition, it is imperative to ensure the optimum level of detail capture for both analysis and processing of data for automating the scoring process. Through the delivery of logs from one developer for ATC21S, it was apparent that the contents of the responses captured should be recorded under several attributes in a well-structured database to optimise the processing time for scoring complex data and to ensure uninterrupted traffic load on the system. Timestamping was found to be essential for logging response data from the assessment tasks. Timed data, along with database indexing, proved to be useful in sequencing user interactions with the task environment. In the current case, database design allowed the capture of user responses only in corresponding seconds. From the accumulated data it was observed that more precise times (i.e. milliseconds) when users respond may often be required to differentiate sequences of actions that occur almost simultaneously. Multiple actions can be recorded as occurring at the same time (in seconds), but actions do occur consecutively and this should be more accurately reflected in the way they are captured and arranged in the database.

Event types described across different tasks should be defined in a uniform method. Consistency in event definition is important for future developers of similar tasks and for understanding the events they represent. In the present context, the assessment tasks were initially designed by different developers. As a result, the language and format used to define the same event – for instance 'chat' – were quite different and had different naming conventions across the various tasks (e.g. 'Send message', 'Type message', 'Enter text' etc.).

Development of interactive tasks and the capacity to automatically score responses is a resource intensive undertaking, even in traditional and well-defined educational domains (Masters 2010). Due consideration should be given to future analysis needs while designing complex assessments of this nature. Emphasis should be given to understanding the intended use of the data to support inferences to the diagnostic richness that can be pertained through interpretation and analysis. This is important, since extension towards more complex data accumulation in less concisely defined educational domains, such as interactive problem solving, may challenge conventional approaches to scaling educational assessment data and may be inadequately handled (Rupp 2002).

While the content of actions can be assessed, assessing the content of chat is currently beyond the limitations of this project. There are some robust automated text analysis programs that analyse large-volume texts – for example, essays, formal open-ended items and reports. One application of these is the Coh-Metrix (Graesser et al. 2004), a computational linguistics tool, which can analyse text for cohesion, language/discourse, and readability. However, the challenges posed to ATC21S by the use of such a tool were too great. To begin with, as the project is international, there are several different language translations involved, which could lead to translation issues within automated text analysis programs. The automated text analysis software would also need to be quite sophisticated to classify the text blocks into the predefined activity type – for example, chat/action/chat. A further difficulty is the quantity and quality of text that may be present within a task's chat box. Students may provide single word answers or low volumes of text, and the type of software available is designed for large quantities of text. The quality of chat is likely to present problems, including grammatical errors, non-standard syntax, abbreviations, and synonyms or 'text-speak' – all of which involve non-standard spelling that would not be recognised by current software designed for more formal language. A key consideration for future deployment is the identification of ways to capture these text data in an understandable coded form or to translate them into a uniform language (such as English) before they are recorded.

References

Arroyo, I., & Woolf, B. P. (2005). *Inferring learning and attitudes from a Bayesian Network of log file data.* Conference paper presented at the Artificial Intelligence in Education: Supporting learning through intelligent and socially informed technology. Amsterdam, The Netherlands.

Bennett, R. E., Jenkins, F., Persky, H., & Weiss, A. (2003). Assessing complex problem solving performances. *Assessment in Education, 10*(3), 347–359.

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

Chung, G. K., de Vries, L. F., Cheak, A. M., Stevens, R. H., & Bewley, W. L. (2002). Cognitive process validation of an online problem solving assessment. *Computers in Human Behaviour, 18*, 669–684.

Fu, W. T. (2001). ACT-PRO action protocol analyzer: A tool for analyzing discrete action protocols. *Behavior Research Methods, Instruments, and Computers, 33*(2), 149–158.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, and Computers, 36*(2), 193–202.

Greiff, S., Wüstenberg, S., & Funke, J. (2012). Dynamic problem solving: A new assessment perspective. *Applied Psychological Measurement, 36*(3), 189–213.

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* (pp. 37–56). Dordrecht: Springer.

Masters, J. (2010). Automated scoring of an interactive geometry item: A proof-of-concept. *Journal of Technology, Learning, and Assessment, 8*(7). Retrieved April 12, 2013, from http://escholarship.bc.edu/jtla/, http://www.jtla.org

Mills, C. N., Potenza, M. T., Fremer, J. J., & Ward, W. C. (2002). *Computer-based testing: Building the foundation for future assessments*. Mahwah: Lawrence Erlbaum Associates.

OECD. (2013). *PISA 2015: Draft collaborative problem solving framework*. http://www.oecd.org/pisa/pisaproducts/Draft%20PISA%202015%20Collaborative%20Problem%20Solving%20Framework%20.pdf. Accessed 7 July 2014

Pelligrino, J., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.

Polya, G. (1945). *How to solve it* (1st ed.). Princeton: Princeton University Press.

Polya, G. (1957). *How to solve it* (2nd ed.). Princeton: Princeton University Press.

Rowe, M. (1972). *Wait time and rewards as instructional variables, their influence in language, logic, and fate control*. Paper presented at the National Association for Research in Science Teaching, Chicago, IL.

Rupp, A. (2002). Feature selection for choosing and assembling measurement models: A building-block-based organisation. *International Journal of Testing, 2*(3/4), 311–360.

Schoenfeld, A. H. (1985). *Mathematical problem solving*. New York: Academic.

Stahl, R. (1990). *Using "think-time" behaviors to promote students' information processing, learning, and on-task participation. An instructional module*. Tempe: Arizona State University.

Tobin, K. (1987). The role of wait time in higher cognitive level learning. *Review of Education Research, 57*(1), 69–95.

Williamson, D. M., Mislevy, R. J., & Bejar, I. I. (2006). *Automated scoring of complex tasks in computer-based testing*. Mahwah: Lawrence Erlbaum Associates.

Zoanetti, N. P. (2010). Interactive computer based assessment tasks: How problem-solving process data can inform instruction. *Australasian Journal of Educational Technology, 26*(5), 585–606.

Chapter 7
Task Characteristics and Calibration

Patrick Griffin, Esther Care, and Susan-Marie Harding

Abstract This chapter outlines the procedures for calibrating and establishing the properties of the collaborative problem solving tasks in the ATC21S™ project (The acronym ATC21S™ has been globally trademarked. For purposes of simplicity the acronym is presented throughout the chapter as ATC21S.). The chapter deals with the interpretation of these tasks and provides an outline of how they were used, discussing the data they yielded, the interpretation of the CPS construct and the calculation of the student skill-levels measured. Using item response theory, the tasks were calibrated separately and jointly. One and two parameter item response models were used to explore the data and to determine dimensionality. The data were analysed on one, two and five dimensions, corresponding with the theoretical components of the collaborative problem solving construct. Tasks were calibrated in sets of three and these sets were used to determine that there were no significant differences between countries in the difficulty of the items. Difference in mean latent ability of Student A and Student B was also analysed, and it was concluded that there was no advantage or disadvantage to students adopting either role. The task calibrations were used to determine the hierarchy of the indicators, and describe student competency levels as measured by the tasks. Skills progressions were created for one, two and five possible dimensions as interpretations of the collaborative problem solving continuum. In this chapter we describe the methods used to develop the progressions from novice to expert, which provide a framework for teachers to use in interpreting their observations of student behaviour regarding collaborative problem solving.

Introduction

Eleven collaborative problem solving (CPS) tasks were developed for the ATC21S project. Tasks are described in detail in Care, Griffin, Scoular, Awwal and Zoanetti (2015; Chap. 3). Two broad classes of task were developed. There was a set of

P. Griffin (✉) • E. Care • S.-M. Harding
Assessment Research Centre, Melbourne Graduate School of Education,
University of Melbourne, Parkville, VIC, Australia
e-mail: p.griffin@unimelb.edu.au

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_7

items based on problem solving tasks originally developed at the University of Melbourne (see Griffin et al. 2013) and a second set developed by World Class Arena in the UK. Tasks in the first set assumed no prior knowledge and were based on the hypothetico-deductive approach described in Hesse, Care, Buder, Sassenberg and Griffin (2015; Chap. 2). Tasks in the second set were curriculum related and, while they could be solved collaboratively without background knowledge, any such knowledge could help in the problem solving process, though not necessarily in the collaboration required.

Hypothetico-deductive

1. Hot chocolate – maximising sales and profits across given regions using recipe and market information
2. Olive oil – an old problem using 5 l and 3 l jugs, each controlled by a different person, to obtain 4 l of oil
3. Sunflower – the mixing of two plant foods to maximise the height of a plant
4. Clowns – identifying how a clown machine functions and comparing this with another clown machine

Curriculum – Science

1. Plant growth – control temperature or light to work out the factors that maximise growth
2. Balance – weights are used to balance a beam
3. Shared garden – symmetrical task to rejuvenate dead plants in a shared garden

Curriculum – Mathematics

1. Game of 20 – the object of the game is to work out the ideal numbers to reach 20 before the computer
2. Warehouse – cameras are placed to cover all available line-of-sight positions to protect boxes that are stored in a warehouse.
3. Hexagons – mathematical rules are identified as number patterns are presented in hexagons
4. Small pyramids – work out the mathematical rule predicting the top number in a series of small pyramids

Concept and Construct Mapping

An early step in the development of these tasks involved concept and construct mapping to identify the elements of collaborative problem solving that were considered important and that could, in all likelihood, influence curriculum change processes. The new or emerging curriculum in the 21st century needs to become a vehicle for changes in teaching methods and one in which students are encouraged to learn in a more proactive manner, due to the interactive nature of the new technologies. The adoption of this curriculum implies that teachers will encourage students to move away from rote learning. There is also an intention to introduce more open, non-content curriculum based 'problem solving' into the curricula (see for example, changes in Hong Kong and Australia). Curriculum as defined by the ATC21S project would emphasise the social skills of perspective taking, participation and social regulation, and the cognitive skills of task regulation and knowledge building. These five strands would be expected to develop throughout elementary education to middle high school levels and beyond.

Specification and Blueprints

The first step in deciding what the assessment tasks would measure was to analyse the content and construct definitions (as outlined by Hesse et al. 2015). A table of specifications (or blueprint) was constructed which determined the selection of learning goals and, in turn, the tasks and source materials.

Table 7.1 lists the elements and behavioural indicators for the social and cognitive strands of collaborative behaviour. Collaborative problem solving was conceived as consisting of social and cognitive dimensions. Within the social component, the strands included participation, perspective taking and social regulation; and within the cognitive component, strands included task regulation or knowledge building. For each of these indicators a series of rubrics of three levels was developed. These were at 'low', 'middle' and 'high' levels of operation. The rubrics were used to inform identification of indicators that could be programmed into the tasks.

Table 7.1 Strands of collaborative problem solving behaviour

Social elements	
Participation	**Indicative behaviour**
Action	Activity within environment
Interaction	Interacting with, prompting and responding to contributions of others
Task completion	Undertaking and completing a task or part of a task individually
Perspective taking	
Adaptive responsiveness	Ignoring, accepting or adapting contributions of others
Audience awareness	Awareness of how to adapt behaviour to increase suitability for others
Social regulation	
Negotiation	Achieving a resolution or reaching compromise
Self-evaluation	Recognising own strengths and weaknesses
Trans active memory	Recognising strengths and weaknesses of others
Responsibility initiative	Assuming responsibility for ensuring parts of the task are completed by the group

(continued)

Table 7.1 (continued)

Cognitive elements	
Task regulation	
Resource management	Managing resources or people to complete a task
Collect elements of information	Explores and understands elements of the task
Systematicity	Implements possible solutions to a problem and monitors progress
Tolerance for ambiguity/tension	Accept ambiguous situations
Organisation (problem analysis)	Analyse and describe the problem in familiar language
Setting goals	Sets a clear goal for a task
Knowledge building	
Knowledge acquisition	Follows the path to gain knowledge
Relationships (representation and formulation)	Identifies patterns and connections between among elements of the knowledge associated with the problem
Rules "if… then"	Uses understanding of cause and effect to
Hypothesis "what if…" (Reflection on Solution)	Adapts reasoning or course of action as problem solved

The Trial Samples

Data from international trial and calibration samples were collected through country trials, and analysed. The task participation numbers by country are reported in Table 7.2. As the table shows, the different countries did not all use the same set of tasks: the Netherlands did not use the Balance or Sunflower tasks, and the United States did not use the Hexagons task.

Task Calibration

The log stream data file for each of the CPS tasks yielded a range of data points. These were coded and scored as indicators of student performance, as described in Adams, Vista, Scoular, Awwal, Griffin, and Care (2015; Chap. 6). The indicators were scored as a mixture of dichotomous and partial credit, the majority of indicators being dichotomous. A total score was tallied for each individual student based upon the sum of the scores associated with indicative behaviours that the student demonstrated.

Initial analyses of each of the tasks identified those indicators that could contribute to a substantive interpretation of the underpinning construct. These were the indicators that had a consistent relationship between the estimate of student ability (total score as a proportion of the maximum possible) and the probability of the individual indicator being demonstrated. This is to say that the underpinning construct was assumed to exist and that the data were explored to identify those indicators that supported this assumption. In order to do this the data were analysed using the one-parameter simple logistic item response model (Rasch (1960/1980). As the Rasch

Table 7.2 Task completions grouped by country

Country	Task											
	Balance	Game of 20	Hot chocolate	Laughing clowns	Plant growth	Small pyramids	Olive oil	Shared garden	Sunflower	Warehouse	Hexagons	Total
Australia	340	256	582	586	316	190	592	202	460	440	166	**4,130**
Costa Rica	22	182	266	276	146	104	466	112	266	248	106	**2,194**
Finland	6	162	190	190	138	148	402	94	308	238	88	**1,964**
Netherlands	0	98	226	226	288	102	224	80	0	244	58	**1,546**
Singapore	258	340	420	396	342	336	344	400	696	530	450	**4,512**
United States	214	210	226	348	134	166	224	152	534	478	0	**2,686**
Total	**840**	**1,248**	**2,000**	**2,022**	**1,364**	**1,046**	**2,028**	**1,040**	**2,264**	**2,178**	**868**	***16,898***

model estimates the probability of success based upon the difference between the task difficulty and the student ability, task difficulty was assumed to be approximated by the frequency of occurrence in the dichotomously scored case. Student ability was assumed to be estimated by proportional score based upon the maximum possible given that not all indicators would be available for all students. There was an additional reason for using the Rasch model. With the response probability set at .5, it was possible to identify the point at which the student ability was equal to the indicator difficulty. When this situation could be described for a group of students and a cluster of indicators, which reflected the elements described in Table 7.1, a substantive interpretation of the stage of competence could be obtained. This in turn makes it possible to provide advice to teachers based on the Vygotsky zone of proximal development.

In calibrating the tasks, the measure of "achievement" or "ability" must be valid across all participating sub-groups. This is important for meaningful interpretations of achievement whether they are at a national or international level. Verifying the variable relies on calibration procedures both within and between student and task samples and this can be achieved with the Rasch simple logistic model (SLM). The simple logistic model has been shown to predict accurately both the behaviour of task indicators and of persons. Other, more complex item response models have consistently generated theoretical and practical difficulties. Given that the purpose is to interpret the underpinning construct, the relationship between student ability and performance on the construct has to remain relatively constant. If the purpose is to explain variability between items and between students, the more parameters that can be entered into the model the better. Hence explaining variation requires a multiparameter model. Identifying the underlying construct requires a Rasch SLM with far more stringent assumptions held.

Interpreting performance on a developmental continuum or construct requires a single parameter model. For example, when guessing and discrimination are used as additional parameters, lengthy computations are required to score the test and the simple one-to-one relationship with the raw score is lost; but these parameters do help to shed light on inter- and intra-national and regional differences in indicator behaviour. However for the main ATC21S study the explanation of differences in achievement and the capacity of the measures to demonstrate growth in student ability formed the main focus. Hence, emphasis was placed on the one-parameter model for calibration and interpretation purposes. What the Rasch SLM imposed, however, was a task design that had a dominant underlying variable which was then operationalised in the student behaviour and interpreted through the activity indicators. It was assumed that this would remain the case even though the tasks were completed by pairs of students and between-student dependence may have existed. That is, the tasks were deliberately constructed to contribute to the interpretation of the construct. Specifications used by the item developers dictated that the major construct underpinned the development. Developers were provided with both the blueprint specifications and the hierarchy of problem solving behaviour that needed to be built into each collaborative task.

For purposes of exploration of the data, each indicator was regarded as equivalent to a test item. The tasks therefore involved complex interrelated steps leading to the collaborative solution of problems. The scoring of the indicators as dichotomous (present/absent was regarded as the equivalent of correct/incorrect) or partial credit enabled the Rasch model to be applied in exploring the relationship between student ability and performance on each of the indicators or items. Given the issue of indicators being present or absent, specific indicators needed to be defined in terms of missing data. Where the indicator was not relevant to a particular task this would be the equivalent of a task or item not being administered. Each task therefore would be considered as the equivalent of a short test, and combinations (or sets) of tasks were used to ensure sufficient data points were available to provide reliable estimates of indicator difficulty and person ability.

According to the SLM, the probability of a given response to an item does *not* depend on which individuals attempt the item but on the pattern of responses given. The model does not depend on which indicators make up the task, nor on the order in which they appear, nor on the responses to preceding indicators in the task. It is assumed that the individual's ability to demonstrate the indicative behaviour is conditioned only by ability and not by motivation, guessing tendency or any personal attribute other than the ability in the domain of interest. The model assumes just one indicator parameter (difficulty – treated as the equivalent of frequency of occurrence such that low frequency is regarded as difficult and high frequency is regarded as easy) and a single person parameter (ability). The difficulty and ability parameter estimates were mapped onto a single interval scale. Both parameters were measured in the same units, called logits. The single scale enabled both persons and indicators to be placed on the same continuum defining an underlying variable, which was interpreted in terms of the skills required for the student to provide an observable indicator response.

Fit to the Model

Two measures of accuracy of the test procedure were used. The first was the measure of the error of measurement for each of the item difficulty estimates and student ability estimate. The second was a measure of the extent to which the data were consistent with the requirements of the Rasch model. This measure is the mean squared differences between the estimated (or modelled) difficulty and the observed difficulty of each score point, weighted by the variance of the assigned scores. This is called the Infit mean square (Information Weighted Mean Squared residual goodness of fit statistic). The expected value of the Infit is 1.0 and accepted range of these values lies between 0.77 and 1.30 (Adams and Khoo 1993) and when the items sets are all within these limits, this is taken as evidence of a single dominant dimension underpinning the test performances of the students.

Fit is useful for investigating how accurately the model can be used to predict performance. The relationship between ability and performance should be such that as ability increases, the chances of demonstrating the indicative CPS behaviour on each task also increases. When the relationship between ability (or difficulty) and performance breaks down, the fit statistic indicates the extent to which the relationship has been lost. In the case of the ATC21S indicators, all fit statistics were excellent, as will be shown.

Reliability Estimates

Traditional approaches to reliability estimation assume a classical measurement model. In that approach it is assumed that the raw score is composed of two components, a true score and an error component. Cronbach's approach is to calculate the ratio of the true score variance to the total variance and this is classically known as reliability or Cronbach alpha. Both Cronbach and Rasch Separation Indices are estimates of the ratio of 'true' measure variance to the 'observed' measure variance. A Rasch estimate of reliability allows an investigation of two measures. The first is the student variance with measurement error removed and the second is the average precision of the student measures. The ratio of the adjusted standard deviation (with the error removed) to the average precision (that is the mean measure standard error) is called the separation index. In the case of the Rasch reliability a separation index can also be devised and the interpretation of this has been discussed by Wright and Masters (1982) who showed that the item separation index can be used as an index of construct validity and the person separation index can be used as an index of criterion validity.

Establishing Validity Using Rasch Modelling

Wright and Masters (1982) showed that separating the items and identifying the skills underpinning each item could help define the variable underpinning the test. Items that cluster together do not provide sufficient information about the variable to allow interpretation, but if a sequence of clusters can be identified and each has a cohesive and interpretable meaning the variable can be clearly identified. Their emphasis on clusters of items is often missed when observers focus on each individual item or indicator and its position on the construct or variable map. The importance of interpreting clusters of items cannot be overemphasised. Once clusters of items have been calibrated along the variable, they can be interpreted in terms of the task developers' intentions. To achieve this, a skills audit of each of the indicators has to be undertaken. However, even the task developers' intention can sometimes be misleading and a pilot study with students from the target population is often used to identify the cognitive skills used by students obtaining the correct answer. Examination of the indicator score threshold locations provides information about the connections between an indicator and the underlying construct the set of indicators was expected to measure. In addition to providing a 'map' of students' increasing understanding, examination of model fit can provide information about how much it is justified to interpret and measure the underlying construct with the particular set of indicators chosen (Wilson and Adams 1995). Good fit to the model suggests that the student performances on the items (indicators) are measuring the same single dimension construct, that is, the assessment can be argued to have demonstrated construct validity.

Calibration of the Tasks and Interpretation of the Data

The collaborative problem solving tasks are completed by Student A and Student B, who are expected to collaborate and solve the problem. To assess the ability of each student to collaborate, we have used some indicators which are specific to each student, some indicators which are common to both students (common indicators) and some indicators which are used for all tasks (global indicators). Presented in the following section are the calibration statistics and psychometric properties of four tasks as examples. Again, the tasks are not used on their own to calculate students' ability; a set of tasks is required. The calibration of each task individually provides information we can use to identify which tasks should be used together as a set. The difficulty of the task overall, and its spread of indicators and elements from the conceptual framework, are important factors for deciding which tasks should be taken together. The statistics for each task alone are fairly robust, but the psychometric properties improve greatly when more data is obtained from each student by completion of more than one task.

Olive Oil

The calibration estimates for the olive oil task indicators have been presented in Table 7.3. For each item, the summary statistics were as follows. The item difficulty (logit) is reported and measurement error (SE) for each difficulty, the weighted fit

Table 7.3 Olive oil parameter estimates and fit

Variables				Unweighted fit				Weighted fit			
Item		Estimate	Error	MNSQ	Confidence interval		T	MNSQ	Confidence interval		T
1	1	0.032	0.047	0.71	0.91	1.09	−7.3	0.81	0.93	1.07	−5.9
2	2	−0.349	0.052	1.07	0.86	1.14	1	1.04	0.92	1.08	1
3	3	3.058	0.057	1.27	0.85	1.15	3.3	1.14	0.69	1.31	0.9
4	4	−1.687	0.047	0.99	0.91	1.09	−0.1	1.03	0.93	1.07	0.8
5	5	−0.154	0.05	1.28	0.89	1.11	4.7	1.24	0.93	1.07	5.9
6	6	−0.692	0.049	1.13	0.89	1.11	2.3	1.08	0.93	1.07	2.4
7	7	0.403	0.048	0.73	0.91	1.09	−6.7	0.82	0.92	1.08	−4.8
8	8	−0.922	0.055	1.21	0.83	1.17	2.4	1.15	0.85	1.15	2
9	9	0.66	0.049	0.55	0.91	1.09	−12.2	0.72	0.91	1.09	−7
10	10	2.801	0.057	1.38	0.82	1.18	3.7	1.09	0.72	1.28	0.7
11	11	1.202	0.056	1.24	0.82	1.18	2.5	1.09	0.82	1.18	1
12	12	0.717	0.049	0.68	0.91	1.09	−8.1	0.8	0.91	1.09	−4.7
13	13	0.696	0.048	1.05	0.91	1.09	1.1	1.04	0.92	1.08	0.9
14	14	1.428	0.047	1.7	0.93	1.07	16.2	1.23	0.91	1.09	4.6
15	15	−4.731	0.055	1.09	0.93	1.07	2.5	1	0.77	1.23	0.1
16	16	*−2.463*	0.198	1.04	0.94	1.06	1.2	1	0.94	1.06	0.1

(infit) and un-weighted fit (outfit) estimates and the confidence intervals of these estimates are included along with the t value. The weighted or infit statistic is most valuable to us as this measure is sensitive to the patterns of responses to items targeted to the person. The un-weighted fit is sensitive to unexpected responses to the items with difficulty very different from the person ability (outliers). The mean weighted fit (INFIT) was 1.01 with a variance of 0.022, indicating good fit to the model. All of the infit values of the final set of indicators were within the range of 0.7–1.3 and hence there was evidence of a dominant underlying dimension in the variable being measured. This evidence indicates that the test was measuring a single dominant variable and that a single dominant latent variable underpinned the set of items. It also indicated that the test successfully separated the students on the basis of ability (i.e. that it possessed acceptable criterion validity) as well as demonstrating construct validity. On the latter point however, there is no external evidence of the nature of the construct criterion.

The olive oil task was slightly difficult for the students in that the mean student ability measures were −0.865 logits compared to a logit mean ability of 0 the indicator mean difficulty. There were 2,028 students who completed this task.

The person separation reliability index was 0.562. This is acceptable as there are few data points in any one task to have any higher reliability of person separation. Many of the characteristics of the task can be identified from the variable map, which has been presented shown for the olive oil task as Fig. 7.1. The chart has several sections to it. Working from the left of the figure the first characteristic of the chart is a scale that ranges from −5.0 to +4.0. This is the logit scale and is the metric of the Rasch analysis that enables student ability and item difficulty to be mapped onto the same scale. The distribution of student ability is presented next and each 'X' represents approximately 3.1 students.

Laughing Clowns

The calibration estimates for the Clowns task indicators have been presented in Table 7.4. As with the olive oil task, the summary statistics are described. The mean weighted fit (INFIT) was 1.00 with a variance of 0.0004, indicating excellent fit to the model. All of the infit values of the final set of indicators were well within the range of 0.7–1.3; again this indicates that the test successfully separated the students on the basis of ability. The Laughing Clowns task was relatively easy for the students in that the mean student ability measures were 0.105 logits with a variance of 0.669.

The variance of item difficulty levels was 1.58 showing that the range of indicator difficulty was greater than the spread of student ability. Given this situation, the reliability of item separation of 0.999 was exceptionally high and according to Wright and Masters (1982).This index can be used as evidence of construct validity – the extent to which the indicators enable an interpretation of the underlying construct. There were 2,022 students who completed this task.

```
  Logits                 Students                            Indicators
---------------------------------------------------------------------------------------
   4                                          |                                       |
                                              |                                       |
                                              |                                       |
                                              |                                       |
                                              |                                       |
                                              |                                       |
   3                                          |3                                      |
                                              |10                                     |
                                             X|                                       |
                                             X|                                       |
                                              |                                       |
                                            XX|                                       |
   2                                       XXX|                                       |
                                          XXXX|                                       |
                                        XXXXXX|                                       |
                                         XXXXX|                                       |
                                       XXXXXXX|14                                     |
                                      XXXXXXXX|11                                     |
   1                                XXXXXXXXXX|                                       |
                               XXXXXXXXXXXXXXX|                                       |
                         XXXXXXXXXXXXXXXXXXXXX|9 12 13                                |
                            XXXXXXXXXXXXXXXXXX|                                       |
                             XXXXXXXXXXXXXXXXX|7                                      |
                         XXXXXXXXXXXXXXXXXXXXX|                                       |
   0                XXXXXXXXXXXXXXXXXXXXXXXXXX|1                                      |
                   XXXXXXXXXXXXXXXXXXXXXXX|5                                      |
               XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|2                                      |
                   XXXXXXXXXXXXXXXXXXXXXXXXXXX|                                       |
          XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|6                                      |
          XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                       |
  -1  XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|8                                      |
             XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                       |
            XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                       |
                    XXXXXXXXXXXXXXXXXXXXXXXXXX|                                       |
                 XXXXXXXXXXXXXXXXXXXXXXXXXXXXX|4                                      |
                    XXXXXXXXXXXXXXXXXXXXXXXXXX|                                       |
  -2                  XXXXXXXXXXXXXXXXXXXXXXXX|                                       |
                         XXXXXXXXXXXXXXXXXXXXX|                                       |
                            XXXXXXXXXXXXXXXXXX|                                       |
                             XXXXXXXXXXXXXXXXX|16                                     |
                                 XXXXXXXXXXXXX|                                       |
                                  XXXXXXXXXXXX|                                       |
  -3                                  XXXXXXXX|                                       |
                                       XXXXXXX|                                       |
                                          XXXX|                                       |
                                           XXX|                                       |
                                           XXX|                                       |
                                            XX|                                       |
  -4                                        XX|                                       |
                                              |                                       |
                                             X|                                       |
                                              |                                       |
                                              |                                       |
                                              |15                                     |
  -5                                         X|                                       |
                                              |                                       |
                                             X|                                       |
                                              |                                       |
                                              |                                       |
                                              |                                       |
=======================================================================================
```

Fig. 7.1 Olive oil variable map. Each 'X' represents 3.1 cases

Table 7.4 Clowns parameter estimates and fit

Variables				Unweighted fit				Weighted fit			
Item		Estimate	Error^	MNSQ	Confidence interval		T	MNSQ	Confidence interval		T
1	1	3.106	0.046	1.13	0.91	1.09	2.8	1.02	0.75	1.25	0.2
2	2	−0.686	0.04	0.97	0.91	1.09	−0.5	0.98	0.95	1.05	−0.7
3	3	1.01	0.04	1	0.91	1.09	−0.1	1	0.94	1.06	0
4	4	0.454	0.039	1	0.91	1.09	−0.1	1	0.97	1.03	−0.2
5	5	−0.435	0.039	1	0.91	1.09	0	1	0.96	1.04	−0.1
6	6	−1.409	0.042	0.98	0.91	1.09	−0.5	0.99	0.9	1.1	−0.2
7	7	−0.218	0.039	1.01	0.91	1.09	0.3	1.01	0.97	1.03	0.9
8	8	0.895	0.039	0.98	0.91	1.09	−0.4	0.99	0.95	1.05	−0.5
9	9	0.267	0.039	1.06	0.91	1.09	1.3	1.06	0.98	1.02	4.5
10	10	−0.657	0.04	1	0.91	1.09	0	1	0.95	1.05	0
11	11	1.094	0.04	0.98	0.91	1.09	−0.5	0.99	0.94	1.06	−0.5
12	12	0.424	0.039	1.05	0.91	1.09	1	1.04	0.97	1.03	2.8
13	13	−0.523	0.039	0.98	0.91	1.09	−0.5	0.98	0.96	1.04	−0.9
14	14	−1.416	0.042	0.97	0.91	1.09	−0.7	0.98	0.9	1.1	−0.3
15	15	−0.011	0.039	0.99	0.91	1.09	−0.2	0.99	0.98	1.02	−1
16	16	1.464	0.039	1.04	0.92	1.08	1	1.02	0.93	1.07	0.5
17	17	−2.714	0.045	0.94	0.92	1.08	−1.6	0.99	0.8	1.2	−0.1
18	18	*−0.646*	0.166	0.98	0.94	1.06	−0.7	0.98	0.97	1.03	−0.9

The person separation reliability index was only 0.197. This demonstrates that this task could not be used alone for estimating ability parameters of the students who completed the task. Wright and Masters (1982) use the person separation index as a measure of criterion validity. The variable map is presented in Fig. 7.2.

Balance

The difficulty estimates for the Balance task indicators are presented in Table 7.5. As with other tasks, the summary statistics are presented. The mean weighted fit (INFIT) of the indicators was 1.00 with a variance of 0.023, a good fit to the model. The mean student ability estimate was −0.156 with a variance of 3.05. There were 840 students who completed this task.

The person separation reliability index was 0.606 suggesting that the balance task would help to differentiate students on the basis of estimated abilities. The variable map is presented in Fig. 7.3.

```
Logits                 Students                                   Indicators
 --------------------------------------------------------------------------------------
                                               |1                                     |
                                               |                                      |
                                               |16                                    |
                                               |                                      |
                                               |                                      |
                                               |                                      |
                                              X|                                      |
                                               |                                      |
                                              X|                                      |
                                              X|11                                    |
                                             XX|                                      |
    1                                        XX|3                                     |
                                            XXX|                                      |
                                              X|8                                     |
                                          XXXXX|                                      |
                                          XXXXX|                                      |
                                        XXXXXXX|                                      |
                                     XXXXXXXXXX|                                      |
                                  XXXXXXXXXXXXX|                                      |
                              XXXXXXXXXXXXXXXXX|                                      |
                                XXXXXXXXXXXXXXX|                                      |
                           XXXXXXXXXXXXXXXXXXXX|                                      |
                            XXXXXXXXXXXXXXXXXXX|4                                     |
                     XXXXXXXXXXXXXXXXXXXXXXXXXX|12                                    |
                XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                      |
            XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                      |
                   XXXXXXXXXXXXXXXXXXXXXXXXXXXX|9                                     |
           XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                      |
         XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                      |
            XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                      |
       XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                      |
            XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                      |
    0         XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|15                                    |
           XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                      |
                    XXXXXXXXXXXXXXXXXXXXXXXXXXX|                                      |
                    XXXXXXXXXXXXXXXXXXXXXXXXXXX|                                      |
                        XXXXXXXXXXXXXXXXXXXXXXX|7                                     |
                            XXXXXXXXXXXXXXXXXXX|                                      |
                            XXXXXXXXXXXXXXXXXXX|                                      |
                        XXXXXXXXXXXXXXXXXXXXXXX|                                      |
                                 XXXXXXXXXXXXXX|                                      |
                                  XXXXXXXXXXXXX|5                                     |
                                     XXXXXXXXXX|13                                    |
                                      XXXXXXXXX|                                      |
                                         XXXXXX|                                      |
                                           XXXX|10 18                                 |
                                          XXXXX|2                                     |
                                            XXX|                                      |
                                               |                                      |
                                              X|                                      |
                                               |                                      |
                                               |                                      |
   -1                                          |                                      |
                                               |                                      |
                                               |                                      |
                                               |                                      |
                                               |                                      |
                                               |                                      |
                                               |                                      |
                                               |6 14 17                               |
 ======================================================================================
```

Fig. 7.2 Laughing clowns variable map. Each 'X' represents 2.9 cases

Table 7.5 Balance parameter estimates and fit

Variables				Unweighted fit				Weighted fit			
Item		Estimate	Error^	MNSQ	Confidence interval		T	MNSQ	Confidence interval		T
1	1	−1.796	0.076	0.57	0.86	1.14	−7.3	0.74	0.85	1.15	−3.7
2	2	2.174	0.088	1.16	0.7	1.3	1.1	1.07	0.7	1.3	0.5
3	3	−1.412	0.075	0.68	0.86	1.14	−5.2	0.77	0.87	1.13	−3.6
4	4	−1.357	0.074	0.9	0.86	1.14	−1.5	0.97	0.87	1.13	−0.5
5	5	−1.045	0.076	1.07	0.85	1.15	1	1.04	0.87	1.13	0.6
6	6	3.341	0.085	0.96	0.86	1.14	−0.6	1.05	0.67	1.33	0.3
7	7	−0.029	0.091	1.26	0.16	1.84	0.7	1.25	0.47	1.53	0.8
8	8	−2.441	0.082	0.45	0.85	1.15	−9.1	0.73	0.76	1.24	−2.4
9	9	2.758	0.082	0.83	0.86	1.14	−2.5	1.01	0.76	1.24	0.1
10	10	2.692	0.082	1.25	0.86	1.14	3.4	1.03	0.77	1.23	0.3
11	11	−1.296	0.078	0.84	0.85	1.15	−2.2	0.91	0.85	1.15	−1.2
12	12	2.241	0.08	0.92	0.85	1.15	−1	0.99	0.81	1.19	−0.1
13	13	1.941	0.072	1.9	0.89	1.11	12.3	1.22	0.88	1.12	3.4
14	14	−4.266	0.087	1.27	0.89	1.11	4.4	1.15	0.56	1.44	0.7
15	15	*−1.506*	0.302	1.16	0.9	1.1	3	1.09	0.9	1.1	1.8

Game of 20

The difficulty estimates for the ‘Game of 20’ indicators are presented in Table 7.6. The mean weighted fit (INFIT) was 0.992 with a variance of 0.002, indicating a good fit to the model and a capacity to differentiate among the students on their ability estimates. However ‘Game of 20’ was relatively difficult for the students in that the mean student ability measures was −1.613 logits with a variance of 1.58. The person separation reliability index was 0.552. The variance of indicator difficulty estimates was 3.80 with an item separation reliability index of 0.999. There were 1,248 students who completed this task. Game of 20 would need to be administered together with other tasks to obtain reliable estimates of student ability (Fig. 7.4).

The discussion of individual tasks has shown a good fit to the model in that the tasks appear to provide ability and difficulty estimates consistent with modelled parameters. The main point is that no one task on its own can provide estimates of person ability with the accuracy required for identifying efficient and effective points of teaching intervention and certainly not for policy development, as discussed by Adamson and Darling-Hammond (2015; Chap. 15). This should also provide a cautionary note to developers and users in other settings or alliances that scoring protocols need to be validated for each task. Reliance on a few data points as indicators of collaboration or other characteristics are not likely to provide accurate evidence of development. It is clear from the above discussion that tasks need to be ‘bundled’ into sets of tests in order to provide accurate estimates of student ability, and that the data point efficiency of tasks needs to be taken seriously. The following section illustrates the gains in accuracy, validity and efficiency when sets of tasks are jointly calibrated.

```
Logits                 Students                            Indicators
 ---------------------------------------------------------------------------------------
                                               |                                      |
                                               |                                      |
                                               |                                      |
    5                                          |                                      |
                                               |                                      |
                                               |                                      |
                                               |                                      |
                                               |                                      |
    4                                          |                                      |
                                              X|                                      |
                                              X|                                      |
                                             XX|6                                     |
                                             XX|                                      |
    3                                       XXX|                                      |
                                           XXXX|9                                     |
                                            XXX|10                                    |
                                          XXXXX|                                      |
                                       XXXXXXXX|2 12                                  |
    2                            XXXXXXXXXXXXXX|                                      |
                             XXXXXXXXXXXXXXXXXX|13                                    |
                               XXXXXXXXXXXXXXXX|                                      |
                          XXXXXXXXXXXXXXXXXXXXX|                                      |
               XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                      |
                  XXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                      |
    1               XXXXXXXXXXXXXXXXXXXXXXXXXXX|                                      |
        XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                      |
              XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                      |
        XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                      |
               XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                      |
    0       XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|7                                     |
              XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                      |
       XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                      |
          XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                      |
                    XXXXXXXXXXXXXXXXXXXXXXXXXXX|                                      |
   -1             XXXXXXXXXXXXXXXXXXXXXXXXXXXXX|5                                     |
               XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|11                                    |
                          XXXXXXXXXXXXXXXXXXXXX|3 4 15                                |
                              XXXXXXXXXXXXXXXXX|                                      |
                              XXXXXXXXXXXXXXXXX|1                                     |
   -2                        XXXXXXXXXXXXXXXXXX|                                      |
                                   XXXXXXXXXXXX|                                      |
                                  XXXXXXXXXXXXX|8                                     |
                                   XXXXXXXXXXXX|                                      |
                                     XXXXXXXXXX|                                      |
   -3                                     XXXXX|                                      |
                                       XXXXXXXX|                                      |
                                        XXXXXXX|                                      |
                                         XXXXXX|                                      |
                                              X|                                      |
   -4                                          |                                      |
                                              X|                                      |
                                               |14                                    |
                                               |                                      |
                                               |                                      |
                                              X|                                      |
   -5                                          |                                      |
                                               |                                      |
                                               |                                      |
                                               |                                      |
                                               |                                      |
 =======================================================================================
```

Fig. 7.3 Balance variable map. Each 'X' represents 1.2 cases

Table 7.6 Game of 20 parameter estimates and fit

Variables				Unweighted fit				Weighted fit			
Item		Estimate	Error^	MNSQ	Confidence interval		T	MNSQ	Confidence interval		T
1	1	−2.268	0.056	0.96	0.89	1.11	−0.6	0.97	0.93	1.07	−0.9
2	2	−2.282	0.056	1.01	0.89	1.11	0.3	1	0.93	1.07	0
3	3	−0.773	0.056	0.89	0.89	1.11	−2.1	0.95	0.93	1.07	−1.4
4	4	−0.788	0.056	0.88	0.89	1.11	−2.1	0.94	0.93	1.07	−1.5
5	5	−1.787	0.055	1.04	0.89	1.11	0.7	1.03	0.94	1.06	1
6	6	−1.802	0.055	1.05	0.89	1.11	0.8	1.02	0.94	1.06	0.6
7	7	1.684	0.067	0.94	0.88	1.12	−1	0.99	0.66	1.34	0
8	8	1.511	0.066	0.93	0.88	1.12	−1.2	0.97	0.69	1.31	−0.2
9	9	2.942	0.069	0.63	0.88	1.12	−6.8	0.99	0.33	1.67	0.1
10	10	0.808	0.064	0.84	0.88	1.12	−2.7	0.93	0.79	1.21	−0.7
11	11	0.521	0.063	0.8	0.88	1.12	−3.4	0.92	0.83	1.17	−0.9
12	12	0.764	0.064	0.73	0.88	1.12	−4.8	0.9	0.8	1.2	−1
13	13	−0.786	0.058	1.02	0.88	1.12	0.3	1	0.92	1.08	−0.1
14	14	−0.556	0.059	1.03	0.88	1.12	0.5	1	0.91	1.09	0.1
15	15	1.411	0.066	0.99	0.87	1.13	−0.2	0.98	0.7	1.3	−0.1
16	16	1.194	0.066	0.83	0.87	1.13	−2.8	0.96	0.73	1.27	−0.2
17	17	3.184	0.07	0.39	0.87	1.13	−12.3	0.99	0.21	1.79	0.1
18	18	3.16	0.07	0.41	0.87	1.13	−11.5	0.98	0.22	1.78	0.1
19	19	1.176	0.066	0.72	0.87	1.13	−4.7	0.93	0.74	1.26	−0.5
20	20	1.36	0.066	0.8	0.87	1.13	−3.3	0.93	0.71	1.29	−0.4
21	21	−0.575	0.059	1.01	0.87	1.13	0.1	0.99	0.91	1.09	−0.1
22	22	−0.396	0.06	1.08	0.87	1.13	1.2	1.01	0.9	1.1	0.3
23	23	−2.451	0.06	1.1	0.86	1.14	1.4	1.05	0.9	1.1	1
24	24	−0.942	0.059	1.07	0.86	1.14	1	1.04	0.92	1.08	1
25	25	1.652	0.067	1.12	0.86	1.14	1.6	1.04	0.64	1.36	0.3
26	26	0.658	0.064	1.27	0.86	1.14	3.5	1.06	0.8	1.2	0.6
27	27	2.283	0.069	1.19	0.86	1.14	2.5	1.03	0.47	1.53	0.2
28	28	−0.511	0.061	1.08	0.86	1.14	1.1	1.02	0.89	1.11	0.5
29	29	0.228	0.056	1.29	0.91	1.09	5.7	1.11	0.9	1.1	2.1
30	30	−5.661	0.067	0.82	0.91	1.09	−4.3	0.98	0.6	1.4	−0.1
31	31	*−2.957*	0.342	1.11	0.92	1.08	2.6	1.04	0.93	1.07	1.1

Joint Calibration

Ability and Difficulty Estimates

The estimates of student ability and indicator difficulty were based on a range of approaches to coding, scoring and analyses. These are outlined in Adams et al. (2015). The process focuses on a series of rubrics which were developed and linked to the theoretical construct or conceptual framework for collaborative problem

```
--------------------------------------------------------------------------------------
                                              |9 17 18 27                             |
                                              |                                       |
                                              |7 25                                   |
                                              |8                                      |
                                              |15                                     |
                                              |20                                     |
                                             X|16 19                                  |
                                             X|                                       |
  1                                          X|                                       |
                                             X|10                                     |
                                             X|12 26                                  |
                                            XX|                                       |
                                           XXX|11                                     |
                                           XXX|                                       |
                                           XXX|29                                     |
                                         XXXXX|                                       |
  0                                    XXXXXXX|                                       |
                                      XXXXXXXX|                                       |
                                  XXXXXXXXXXXX|                                       |
                                  XXXXXXXXXXXX|22                                     |
                                      XXXXXXXX|28                                     |
                           XXXXXXXXXXXXXXXXXXX|14 21                                  |
                        XXXXXXXXXXXXXXXXXXXXXX|3                                      |
                       XXXXXXXXXXXXXXXXXXXXXXX|4 13                                   |
 -1          XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|24                                     |
          XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                       |
                       XXXXXXXXXXXXXXXXXXXXXXX|                                       |
      XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                       |
        XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                       |
         XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                       |
              XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                       |
                  XXXXXXXXXXXXXXXXXXXXXXXXXXXX|5 6                                    |
               XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                       |
 -2               XXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                       |
               XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                       |
                   XXXXXXXXXXXXXXXXXXXXXXXXXXX|1 2                                    |
                     XXXXXXXXXXXXXXXXXXXXXXXXX|                                       |
                           XXXXXXXXXXXXXXXXXXX|23                                     |
                         XXXXXXXXXXXXXXXXXXXXX|                                       |
                                   XXXXXXXXXXX|                                       |
                              XXXXXXXXXXXXXXXX|                                       |
 -3                              XXXXXXXXXXXXX|31                                     |
                                     XXXXXXXXX|                                       |
                                      XXXXXXXX|                                       |
                                          XXXX|                                       |
                                         XXXXX|                                       |
                                           XXX|                                       |
                                            XX|                                       |
                                             X|                                       |
                                            XX|                                       |
 -4                                          X|                                       |
                                             X|                                       |
                                              |                                       |
                                              |                                       |
                                              |                                       |
                                              |                                       |
                                              |                                       |
                                              |                                       |
 -5                                           |                                       |
                                              |30                                     |
======================================================================================
```

Fig. 7.4 Game of 20 variable map. Each 'X' represents 1.9 cases

solving as used in the ATC21S project. In addition to the indicators in the activity log file the students provided responses to a series of questions for peer and self-assessment. In some cases these items were needed to ensure there was a match to the blueprint (Hesse et al. 2015).

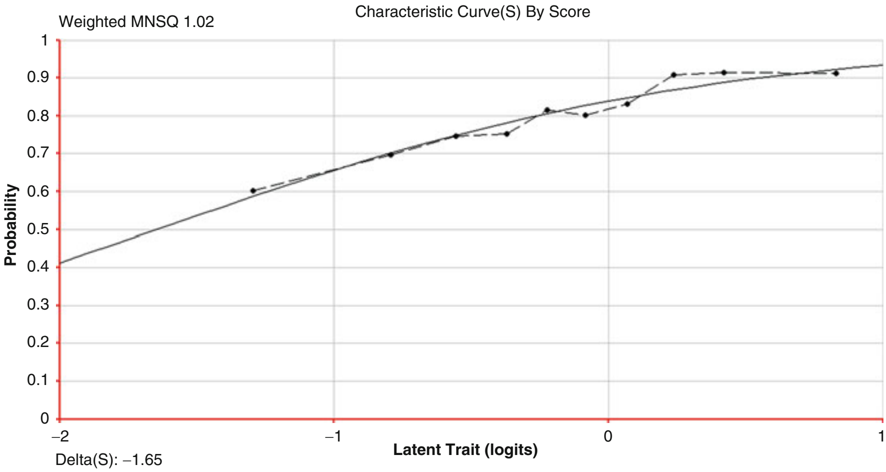


Fig. 7.5 Example item characteristic curve observed (*dots-dashes*) and Modelled (*solid-line* 0)

A one parameter item response model was used to explore the data and to determine the dimensionality of the data. Two dimensions were hypothesised – social and cognitive dimensions. Fit of the model to the data was excellent as described with the one-parameter model used in the task calibrations. An example of fit to the model is illustrated in the item characteristic curve presented in Fig. 7.5 (this particular indicator is U2L009 – consensus of answer for both players). One parameter was preferred for interpretative purposes; two parameters were used for fit improvement. As the fit to the model was excellent in the one parameter model, this model has been used to calibrate the data on the various theoretical dimensions.

One Dimension – One Parameter

The joint calibration combined all data from all students and all tasks in all countries participating in the trials. It was conducted assuming a one-dimensional framework, where there is an assumption that all indicators are measuring a single construct. The separation reliability of the items using a one dimension, one parameter analysis was 0.999, with a person separation reliability statistic of 0.814. These indices are substantially higher than those obtained for single task calibration. The average MNSQ weighted (Infit) was 1.001, with variance 0.002. The mean of latent ability of the students derived from the set of tasks as a whole was –0.182 logits, indicating that all tasks combined resulted in an estimate of student ability slightly lower than the average indicator difficulty – betokening a slightly difficult assessment. There was a good spread of indicator difficulties (Fig. 7.6) and sufficient items at the middle range of student abilities to separate students based on collaborative problem solving ability estimates.

```
Logits                 Students                            Indicators
 -------------------------------------------------------------------------------------
                                               |1 2 5 6 18 20 35 40 42 43 46 52 55 56 |
                                               |58 88 89 91                           |
                                               |47                                    |
    2                                          |31 83                                 |
                                               |27 71 92                              |
                                               |                                      |
                                               |                                      |
                                               |74                                    |
                                               |34                                    |
                                               |3 22 45 48                            |
                                               |7 49                                  |
                                             XX|51 78                                 |
                                             XX|60 117.2                              |
                                             XX|                                      |
                                            XXX|24 37 59 66 82                        |
    1                                    XXXXXX|44 102.2                              |
                                           XXXX|70                                    |
                                     XXXXXXXXXX|                                      |
                                       XXXXXXXX|15 39                                 |
                                XXXXXXXXXXXXXXX|16 26                                 |
                                 XXXXXXXXXXXXXX|                                      |
                             XXXXXXXXXXXXXXXXXX|4 53 100.2 135                        |
                            XXXXXXXXXXXXXXXXXXX|30 36 41 50 65 76 77                  |
                       XXXXXXXXXXXXXXXXXXXXXXXX|33 72 129                             |
                  XXXXXXXXXXXXXXXXXXXXXXXXXXXXX|63 99.2 108.3 115.2                   |
                     XXXXXXXXXXXXXXXXXXXXXXXXXX|96.2 132 144                          |
               XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|10 23 73 85 93.3 101.2 116.2          |
    0        XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|38 75 81 112.2 126                    |
       XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|13 64 69 97.3 106.2 107.2 123 138 153 |
           XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|32 114.2 134                          |
               XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|28 67 90 105.2 111.2 141 147          |
                  XXXXXXXXXXXXXXXXXXXXXXXXXXXXX|14 21 122.2                           |
              XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|62 79 80 98.2 120.2 150               |
                      XXXXXXXXXXXXXXXXXXXXXXXXX|68 113.2 115.1 121.2 137              |
                     XXXXXXXXXXXXXXXXXXXXXXXXXX|17 97.2 116.1 128                     |
                        XXXXXXXXXXXXXXXXXXXXXXX|19 101.1 103.3 104.2                  |
                          XXXXXXXXXXXXXXXXXXXXX|86 95.3 118.3                         |
                             XXXXXXXXXXXXXXXXXX|95.2 106.1 110.3 119.2 131            |
                                XXXXXXXXXXXXXXX|54 57 93.2 103.2 105.1                |
                                XXXXXXXXXXXXXXX|110.2                                 |
   -1                              XXXXXXXXXXXX|95.1 108.2 118.2                      |
                                     XXXXXXXXXX|120.1 121.1 146                       |
                                       XXXXXXXX|25 140 155                            |
                                          XXXXX|87 107.1 125 143 149                  |
                                           XXXX|29 110.1 111.1 122.1                  |
                                            XXX|99.1 100.1 118.1                      |
                                           XXXX|84                                    |
                                             XX|102.1 114.1 119.1                     |
                                               |9 98.1 103.1                          |
                                              X|104.1                                 |
                                            XXX|8 96.1 97.1                           |
                                              X|152                                   |
   -2                                         X|117.1 133                             |
                                             XX|                                      |
                                               |112.1                                 |
                                               |11 136                                |
                                              X|12 93.1 113.1                         |
                                               |                                      |
                                               |145 148 151 154                       |
                                               |61 94 108.1 109 124 127 130 139 142   |
 =====================================================================================
```

Fig. 7.6 One dimension, one parameter variable map. Each 'X' represents 6.6 cases

Two Dimensions – One Parameter

The problem solving (cognitive) aspect and collaborative (social) aspect of each task can be conceptualised as separate dimensions of the construct. As such, separate scaling procedures were conducted wherein these two dimensions were specified to test whether these dimensions are distinct or highly correlated so as to render the assumption of unidimensionality to be tenable. The variable map for two dimension analysis is displayed in Fig. 7.7.

The indicator separation reliability index using a two dimension, one parameter analysis was 0.998 and the person separation reliability was 0.754 on the social domain, and 0.770 on the cognitive domain. The average MNSQ weighted (Infit) was 1.004, with variance 0.003.

The mean latent ability of the students in the social domain was 0.560 logits and in the cognitive domain, −0.824 logits. This indicates that the students performed better in the social domain compared to the cognitive domain. The 2-dimensional scaling showed that the social and cognitive dimensions were highly correlated ($r = 0.788$), with 62 % of student variance shared between the dimensions. This demonstrates that students who have a high ability in the social aspect of collaborative problem solving are more likely to have a higher ability in the cognitive aspect of collaborative problem solving.

Five Dimensions – One Parameter

A model with five strands treated as dimensions was also specified. These five strands (St) are theoretical components of the social dimension (St1 – Participation, St2 – Perspective taking, and St3 – Social regulation) and the cognitive dimension (St4 – Task regulation and St5 – Knowledge building). The variable map for the five dimension model is presented in Fig. 7.8. The rationale for these strands is explained in Hesse et al. (2015).

The item separation reliability index was 0.999 as previously stated and person separation reliability indices for the strands was as follows; social dimension, St1–Participation = 0.671, St2 – Perspective taking = 0.510, and St3 – Social regulation = 0.728, cognitive dimension, St4 – Task regulation = 0.735 and St5 – Knowledge building = 0.653. These reliabilities are acceptable given the ambitious nature of attempting the measure these five theoretical components as separate dimensions. The reliability of the "Perspective taking" strand was the lowest at 0.510; due in part presumably to the difficulty in measuring students' ability in "Ignoring, accepting or adapting contributions of others" (adaptive responsiveness)

```
Logits    Social    Cognitive      Indicators
 --------------------------------------------------------------
                 |         |1 2 5 46 52 88                    |
                 |         |                                  |
                 |         |71 83                             |
                 |         |6 20                              |
                 |         |55                                |
                 |         |18 35 40 58 74                    |
                 |         |89                                |
                 |         |42 91                             |
    2            |         |                                  |
                X|         |                                  |
                X|         |43                                |
                X|         |56                                |
               XX|         |                                  |
              XXX|         |47                                |
              XXX|         |31                                |
            XXXXX|         |27 92 135                         |
          XXXXXXX|         |77                                |
    1     XXXXXXX|         |34 63 129                         |
         XXXXXXXX|         |22 48 85 132 144                  |
        XXXXXXXXX|        X|3 7 45 49 73 126                  |
       XXXXXXXXXX|        X|51 60 78 81 123 138 153           |
         XXXXXXXX|        X|90 102 141 147                    |
         XXXXXXXX|        X|21 24 37 44 59 66 67 82 101       |
        XXXXXXXXX|       XX|62 79 80 106 150                  |
          XXXXXXX|       XX|68 70 100 105                     |
         XXXXXXXX|      XXX|15 39 99 107                      |
    0       XXXXX|     XXXX|16 19 26 86                       |
            XXXXX|    XXXXX|4 54 57 96 97 111                 |
             XXXX|    XXXXX|30 36 41 50 53 65 76 95           |
              XXX| XXXXXXXX|33 72 93 98 110                   |
               XX|  XXXXXXX|25 103 104 108                    |
                X|  XXXXXXX|10 23 87                          |
               XX| XXXXXXXX|29 38 75                          |
                X|  XXXXXXX|13 32 64 69 84 115                |
                X|  XXXXXXX|28 116 134                        |
   -1           X| XXXXXXXX|14                                |
                X| XXXXXXXX|117 137                           |
                 |   XXXXXX|8 17 128                          |
                 |    XXXXX|133                               |
                 |    XXXXX|120 121                           |
                 |    XXXXX|11 114 122 131 136                |
                 |     XXXX|                                  |
                 |      XXX|112 154                           |
                 |       XX|118 119 140 146 155               |
   -2            |       XX|125 143 149                       |
                 |        X|113                               |
                 |        X|61 142 145                        |
                 |        X|9 127 130                         |
                 |        X|                                  |
                 |         |139                               |
                 |         |151 152                           |
                 |         |                                  |
                 |        X|94 124 148                        |
   -3            |         |                                  |
                 |         |12                                |
                 |         |                                  |
                 |         |                                  |
                 |         |109                               |
                 |         |                                  |
                 |         |                                  |
 ==============================================================
```

Fig. 7.7 Two dimension, one parameter variable map. Each 'X' represents 33.1 cases

```
Logits     St1       St2       St3      St4      St5       Indicators
 --------------------------------------------------------------------------------------------
                 |         |         |         |         |                                  |
    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|  XXXXXXX|  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            |         |         |         |         |                                  |
                 |         |         |         |         |                                  |
                 |         |         |         |         |                                  |
                 |         |         |         |         |                                  |
                 |         |         |         |         |                                  |
                 |         |         |         |         |                                  |
 ============================================================================================
```

Fig. 7.8 Five dimension, one parameter variable map. Each ‘X’ represents 45.8 cases

Table 7.7 Correlation matrix

	Strand				
Strand	1	2	3	4	5
Strand 1					
Strand 2	0.565				
Strand 3	0.850	0.332			
Strand 4	0.781	0.549	0.748		
Strand 5	0.703	0.683	0.482	0.693	

Note: Values below the diagonal are correlations

and "Awareness of how to adapt behaviour to increase suitability for others" (audience awareness –mutual modelling).

The mean latent ability estimates for the strands were as follows; social dimension; St1–Participation = 0.368, St2 – Perspective taking = 1.247, and St3 – Social regulation = 0.090, cognitive dimension; St4 – Task regulation = −0.500 and St5 – Knowledge building = −1.159. The correlations of dimensions are shown in Table 7.7. Students scored highest in the Perspective taking strand, although this was arguably the most difficult to measure.

As can be seen in Table 7.7, the highest correlating strands are the Participation and Social regulation strands (0.850) of the social dimension, with 72 % shared variance. The lowest correlation is between Perspective taking and Social regulation. The implications of the differences in students' mean latent abilities on the different theoretically derived stands will be investigated in volume 3 of this series, where results of this study will be analysed in depth.

Sets Analysis

As described previously, the tasks were not designed to be used in isolation. Some tasks provide too few data points for interpretation purposes when considered alone. With the many complexities of collaborative problem-solving skills and different techniques student use, more than one type of task is recommended to be used in an assessment session in order to obtain an accurate description of students' ability.

Tasks were grouped together, with at least three tasks in each set, taking approximately 40–45 min for the students to complete. Compared to a 40 item multiple choice test, a 60 min assessment should yield at least 40 data points to be as efficient. The role taken by students did not affect the estimates of student ability. This should be a generalisable requirement of CPS task design. Students could swap from Student A to Student B between tasks with no consequence. Presented below are the analyses of an example of a set of tasks. A summary table of indicator difficulty estimates, errors of estimates, mean squared weighted and un-weighted fit statistics and confidence intervals are presented. These tables also include estimates for the survey items which were included in each set.

When deciding on which tasks to combine as a set, the following criteria were adhered to:

1. Time taken – where total for a set of tasks should be no more than 40–45 min so they could be realistically completed during a class; this should also yield at least 40 data points;
2. Number and spread of indicators – each sub set should contain indicators with adequate spread to cover all the elements of the conceptual framework and difficulty levels which cover the range of student abilities;
3. Task Difficulty – the overall difficulty of each task was taken into account, with a spread of indicator difficulty estimates closely matched to the expected spread of student ability estimates. This criterion should help to ensure that sets of tasks have a similar overall difficulty;
4. Discrimination values of indicators – the discrimination of the indicators within each task was considered so each set contained an adequate set of indicators that were able to discriminate between students of various CPS abilities.

The following analysis describes the properties of a set of tasks which includes, Olive oil, Laughing clowns and Warehouse. We have chosen this set of tasks as a demonstration, as the individual data from Olive oil and Laughing Clowns have previously been presented. This set also includes the Warehouse task which is a multi-page task based in a mathematical curriculum context.

Notice how the infit, outfit, error and separation reliability statistics are greatly improved by analysing the tasks as a set (Table 7.8 compared to the values reported in Tables 7.3, 7.4, 7.5 and 7.6).

The item separation reliability for Set 4 was 1.000, mean student ability measures were 0.221 logits and person separation reliability was 0.694. The mean weighted Infit of Set 4 was 1.00 with a variance of 0.00 indicating that the tasks successfully separated the students on the basis of ability as well as demonstrating construct validity. The spread of indicator difficulties and student abilities are represented in Fig. 7.9.

Checking for Assessment Bias

Independence of Dyad Partners

There has been some concern and discussion about whether the best way to measure collaborative problem-solving was to engage Human-to-Agent (H2A) or Human-to-Human (H2H) interaction via the Internet. There are reasons to use either or both approaches. In the work of ATC21S project the issue of H2A was not considered although one of the tasks does have the computer playing a game against the student participants.

Table 7.8 Set 4 parameter estimates and fit

Variables				Unweighted fit				Weighted fit			
Item		Estimate	Error^	MNSQ	Confidence interval		T	MNSQ	Confidence interval		T
1	1	0.259	0.041	0.98	0.91	1.09	−0.4	0.98	0.97	1.03	−1.1
2	2	0.471	0.046	0.98	0.87	1.13	−0.3	0.98	0.95	1.05	−0.8
3	3	1.035	0.047	1.07	0.85	1.15	0.9	1.05	0.93	1.07	1.4
4	4	−3.563	0.05	0.89	0.91	1.09	−2.6	1	0.64	1.36	0.1
5	5	0.282	0.041	1.06	0.91	1.09	1.3	1.05	0.97	1.03	2.9
6	6	1.284	0.042	0.92	0.91	1.09	−1.8	0.94	0.93	1.07	−1.8
7	7	−0.576	0.049	1	0.83	1.17	0	1	0.86	1.14	0.1
8	8	0.98	0.041	0.95	0.91	1.09	−1.2	0.95	0.95	1.05	−1.9
9	9	0.049	0.041	1.07	0.91	1.09	1.4	1.06	0.97	1.03	3.1
10	10	1.571	0.048	1.04	0.85	1.15	0.5	1.03	0.87	1.13	0.4
11	11	1.511	0.037	0.93	0.94	1.06	−2.4	0.96	0.94	1.06	−1.5
12	12	1.219	0.036	1.15	0.94	1.06	4.5	1.08	0.96	1.04	3.4
13	13	0.579	0.035	1.11	0.94	1.06	3.5	1.09	0.97	1.03	6.5
14	14	−0.381	0.035	1.03	0.94	1.06	0.8	1.01	0.97	1.03	0.7
15	15	−1.349	0.039	1.01	0.94	1.06	0.5	1.01	0.93	1.07	0.4
16	16	−0.005	0.035	1.11	0.94	1.06	3.5	1.1	0.97	1.03	7.6
17	17	1.021	0.036	1.06	0.94	1.06	1.9	1.04	0.96	1.04	2.1
18	18	1.49	0.042	1.14	0.92	1.08	3.2	1.07	0.93	1.07	1.8
19	19	1.766	0.044	1.12	0.91	1.09	2.6	1.03	0.9	1.1	0.7
20	20	−0.739	0.045	1.11	0.89	1.11	1.9	1.07	0.92	1.08	1.7
21	21	−0.548	0.045	1.04	0.88	1.12	0.6	1.03	0.93	1.07	0.8
22	22	1.261	0.044	1	0.9	1.1	−0.1	1.01	0.93	1.07	0.1
23	23	3.95	0.05	0.75	0.9	1.1	−5.5	0.97	0.61	1.39	−0.1
24	24	2.409	0.046	0.94	0.92	1.08	−1.4	0.99	0.86	1.14	−0.2
25	25	1.623	0.044	0.92	0.91	1.09	−1.8	0.96	0.91	1.09	−0.9
26	26	1.388	0.043	0.94	0.91	1.09	−1.2	0.96	0.92	1.08	−1
27	27	1.596	0.043	1.12	0.91	1.09	2.5	1.05	0.92	1.08	1.1
28	28	1.359	0.044	1.01	0.9	1.1	0.2	1	0.92	1.08	0
29	29	4.071	0.05	0.71	0.9	1.1	−6.5	0.98	0.58	1.42	0
30	30	−0.506	0.017	1.35	0.95	1.05	12.7	1.19	0.94	1.06	6.3
31	31	−0.454	0.025	1.1	0.95	1.05	3.8	1.1	0.96	1.04	4.7
32	32	−0.699	0.024	0.92	0.95	1.05	−3.1	0.93	0.96	1.04	−3
33	33	−0.736	0.019	0.92	0.95	1.05	−3.2	0.94	0.95	1.05	−2.3
34	34	−0.796	0.024	0.89	0.95	1.05	−4.4	0.92	0.95	1.05	−3.4
35	35	−0.161	0.021	1.01	0.95	1.05	0.5	1.01	0.96	1.04	0.5
36	36	−0.79	0.022	1.12	0.95	1.05	4.6	1.1	0.96	1.04	4.3
37	37	−3.814	0.048	0.77	0.95	1.05	−9.7	0.98	0.76	1.24	−0.2
38	38	−0.705	0.019	1.06	0.95	1.05	2.3	1.06	0.94	1.06	1.9
39	39	−0.419	0.023	1	0.95	1.05	0.1	1	0.96	1.04	0
40	40	−1.028	0.027	0.87	0.95	1.05	−5.3	0.9	0.95	1.05	−4.3

(continued)

Table 7.8 (continued)

Variables				Unweighted fit				Weighted fit			
Item		Estimate	Error^	MNSQ	Confidence interval		T	MNSQ	Confidence interval		T
41	41	−0.492	0.024	0.94	0.95	1.05	−2.3	0.95	0.96	1.04	−2.6
42	42	0.14	0.021	1.07	0.95	1.05	2.7	1.06	0.97	1.03	3.5
43	43	0.004	0.028	0.95	0.95	1.05	−2.1	0.95	0.95	1.05	−2
44	44	−0.722	0.02	0.79	0.95	1.05	−8.9	0.86	0.94	1.06	−4.8
45	45	−0.805	0.025	0.82	0.95	1.05	−7.5	0.88	0.95	1.05	−4.8
46	46	−0.338	0.022	0.93	0.95	1.05	−2.7	0.95	0.96	1.04	−2.5
47	47	−3.175	0.048	1.1	0.93	1.07	2.8	1.01	0.76	1.24	0.1
48	48	−1.176	0.038	0.98	0.94	1.06	−0.7	0.99	0.94	1.06	−0.2
49	49	−2.65	0.047	1.02	0.93	1.07	0.5	1	0.8	1.2	0.1
50	50	−0.578	0.036	1.01	0.94	1.06	0.3	1.01	0.96	1.04	0.3
51	51	−3.099	0.048	1	0.93	1.07	−0.1	1.01	0.77	1.23	0.1
52	52	*−1.012*	0.272	1.05	0.94	1.06	1.7	1.03	0.95	1.05	1

The project has always taken the view that H2H interaction is more likely to yield a valid measure of collaboration and that it may be difficult to convince some stakeholders that H2A interaction is collaboration in the real sense of the word. That is there may be a face validity issues. It is clear, however, that a construct may well be defined and clearly delineated as a measurable entity with either approach. The validity of the H2A approach is still to be determined as is its consistency with, or difference from, a construct defined by the analysis of H2H interactions. There has been considerable research using H2A in collaborative learning and it appears to be assumed that this transfers to collaborative problem solving.

The data from the ATC21S project indicates that some of the concerns expressed regarding the efficacy of analyses of H2H interaction may not be well founded. A major concern was the collaboration of students in terms of their different levels of skill in collaboration and the capacity of those students to adjust according to the ability of their partner or partners. The concerns centred on whether asymmetric dyads would affect individual student scores and whether this would be reflected in estimates of population parameters.

The ATC21S has only used dyads. There has been no extension to larger numbers of collaborators. However one of the major questions was whether the difference in the abilities of the collaborators would affect individual and aggregate scores.

Differential Item Functioning

ATC21S has not yet addressed the issue of the effect on individual estimates of difference in abilities between students within a dyad. This still needs to be investigated at an individual and at a classroom level for its implications for teaching.

```
Logits                 Students                            Indicators
 ---------------------------------------------------------------------------------------
                                               |23 29                                 |
                                               |                                      |
                                               |                                      |
                                               |24                                    |
                                               |                                      |
                                               |                                      |
                                               |                                      |
                                               |                                      |
    2                                          |                                      |
                                              X|                                      |
                                               |                                      |
                                             XX|19                                    |
                                             XX|43.2                                  |
                                            XXX|10 25 27                              |
                                           XXXX|11 18                                 |
                                         XXXXXX|                                      |
                                         XXXXXX|26 28                                 |
                                  XXXXXXXXXXXXX|6 22                                  |
                                      XXXXXXXXX|12                                    |
                               XXXXXXXXXXXXXXXX|                                      |
    1                         XXXXXXXXXXXXXXXXX|3 8 17                                |
                             XXXXXXXXXXXXXXXXXX|                                      |
                           XXXXXXXXXXXXXXXXXXXX|                                      |
                       XXXXXXXXXXXXXXXXXXXXXXXX|                                      |
                 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|36.3                                  |
                    XXXXXXXXXXXXXXXXXXXXXXXXXXX|13 31.2                               |
             XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|42.2                                  |
             XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|2                                     |
       XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                      |
            XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|1 5 41.2                              |
              XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|35.2 39.2                             |
                  XXXXXXXXXXXXXXXXXXXXXXXXXXXXX|                                      |
    0     XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX|9 16 46.2                             |
                   XXXXXXXXXXXXXXXXXXXXXXXXXXXX|32.2 40.2                             |
                    XXXXXXXXXXXXXXXXXXXXXXXXXXX|                                      |
                   XXXXXXXXXXXXXXXXXXXXXXXXXXXX|34.2 42.1                             |
                       XXXXXXXXXXXXXXXXXXXXXXXX|30.3 33.3                             |
                           XXXXXXXXXXXXXXXXXXXX|14 30.2 38.3 44.3 45.2                |
                              XXXXXXXXXXXXXXXXX|33.2 35.1                             |
                              XXXXXXXXXXXXXXXXX|7 21 38.2 50                          |
                                 XXXXXXXXXXXXXX|36.2 44.2                             |
                                   XXXXXXXXXXXX|20 30.1 46.1                          |
                                    XXXXXXXXXXX|                                      |
                                         XXXXXX|                                      |
   -1                                      XXXX|39.1 52                               |
                                           XXXX|38.1 44.1                             |
                                            XXX|48                                    |
                                             XX|41.1 45.1                             |
                                               |15 32.1 33.1                          |
                                              X|34.1                                  |
                                             XX|31.1                                  |
                                               |                                      |
                                               |43.1                                  |
                                              X|                                      |
                                               |                                      |
                                               |40.1                                  |
   -2                                         X|                                      |
                                               |                                      |
                                               |                                      |
                                               |4 36.1 37 47 49 51                    |
 =======================================================================================
```

Fig. 7.9 Variable map for a set of tasks acting as a cohesive assessment instrument

The project has addressed the issue however in terms of aggregate data which would be more in line with the purpose of the PISA study. To prevent differential item functioning, any indicator which was asymmetric for Student A and B, was scored as a separate indicator, thus allowing item response theory to take into account any differences.

Table 7.9 Difference in mean latent ability of student A vs. student B for shared garden task

Variables				Unweighted fit				Weighted fit			
Student		Estimate	Error^	MNSQ	Confidence interval		T	MNSQ	Confidence interval		T
1	A	0.018	0.028	0.99	0.88	1.12	–0.1	0.99	0.88	1.12	–0.2
2	B	*–0.018*	0.028	0.99	0.88	1.12	–0.2	1	0.88	1.12	0

Common items where both Student A and Student B are scored are needed for linking the independent or unique indicators to calculate students' overall ability estimate. Initially, some indicators were conceptually designated or assumed to be common for Students A and B. Once adequate trial data had been obtained, frequencies of these indicators for both Student A and B were analysed to determine whether they were acting in the same manner for Student A and B. Only then have we designated these indicators as common items across students.

Differential item function analysis was conducted using Student A and Student B as the subgroups. This was only possible in situations where Student A and B were scored on the same indicator, such situations were more apparent in tasks such as Laughing Clowns or Shared Garden where the task itself was symmetrical and the conceptual basis for which the students were scored was the same. In all tasks, there are some instances in which both students are scored on the same indicator.

Table 7.9 describes the difference in mean latent abilities for the 1,040 students (A and B) completing the Shared Garden task. There was no significant difference between students as was expected. Student A averages were 0.018 logits higher than Student B's, but the measurement error was 0.028 logits. Hence it was concluded that there was no significant difference.

The results of our analyses indicate that, at an aggregate level, it is a matter of indifference whether the dyads are symmetrical or asymmetrical with respect to Student A and Student B role. The overall aggregate position shows little or no difference between Student A and Student B. From this is it is concluded that it is a matter of indifference which student is A and which is B.

Differential item function analysis was performed on indicators completed by both Student A and Student B. Figure 7.10 illustrates an example of the differential item functioning analysis. This is U4L006 of Shared Garden – solution (places the plant in correct position in shared garden). As expected there was no difference in the way these types of items performed when Student A and Student B are compared on an aggregate level.

Task Sets by Country

The task sets were assembled as tools to estimate students' ability in collaborative problem solving. These sets of tasks were designed to be used across countries, so it was important that there was no substantial difference between the difficulties of individual indicators from country to country. Analysis of the data generated by the six participating countries on four different combinations of tasks indicated that

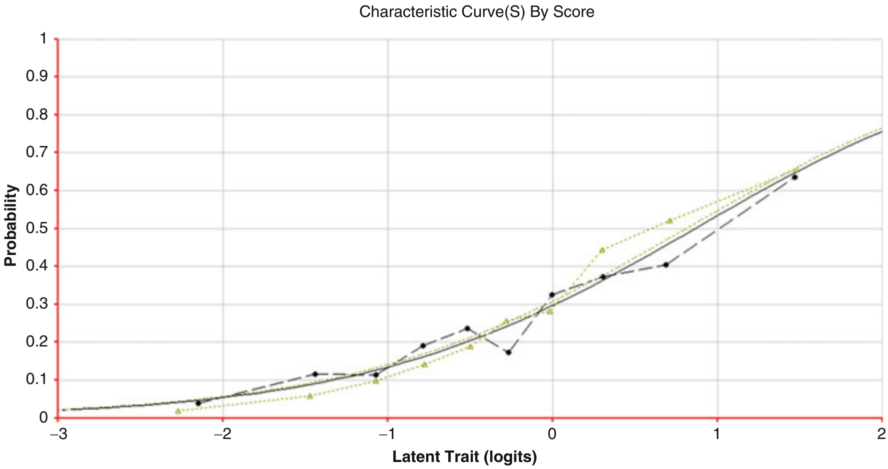


Fig. 7.10 Differential item functioning for student A and student B - typical solution. Notes. Student A observed probabilities black dots • Student B observed probabilities triangles Δ Modelled curve – smooth *solid line*

there was no bias in terms of the difficulties of the indicators across countries. Importantly, these data do not suggest that all countries have students with equal ability levels, but that the indicators themselves are robust enough for measurement in all of the six countries participating in the trials.

Figure 7.11 illustrates the stability of indicator difficulties across countries where they were used. These scatterplots compare psychometric properties of indicators of the set that consisted of Olive Oil, Warehouse and Clowns. Australian data was used as a baseline for comparisons with all other trial countries. The same procedure was applied for all other countries, and for all other sets of tasks with the same results. Confidence bands (95 %) are used to identify any indicator that is significantly different from those established with Australian Data. As can be seen in Fig. 7.11 there are only rare examples of significantly different indicator difficulty estimates. Hence the claim is made that the indicators and their relative difficulties can be used to make a consistent interpretation of the underlying construct in each of the participating countries. In other words the set of tasks and their relevant indicators are argued to be measuring the same thing in each of the six countries independent of the differences of curriculum, language, or culture.

All Task Indicator Difficulties Compared with Australian Estimates

The scatterplots indicate that the set of tasks is robust across countries. We also analysed the pooled data from all indicators. The following scatterplots illustrate the indicator difficulty parameter estimates in Australian trials compared with those of the other five countries tested (Fig. 7.12). These scatterplots indicate that there is a

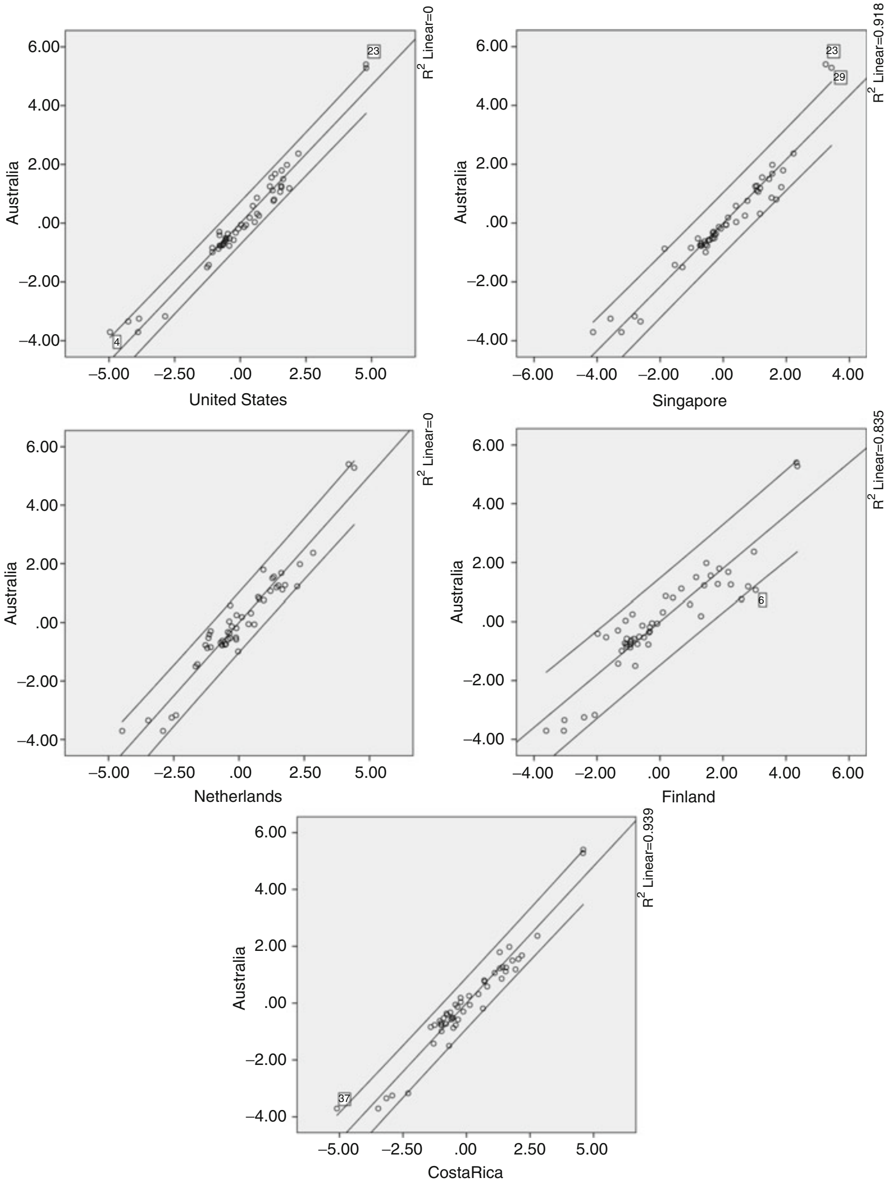


Fig. 7.11 Scatterplot of item delta differences between countries completing the Olive Oil, Clowns and Warehouse tasks as a set

strong correlation (all greater than 0.8) between the difficulty estimates of all indicators when analysed by country. These data support the proposition that the tasks are not only robust when taken in subsets, but that the tasks overall can be used for interpretation of the latent trait in all six countries tested. The 95 % confidence

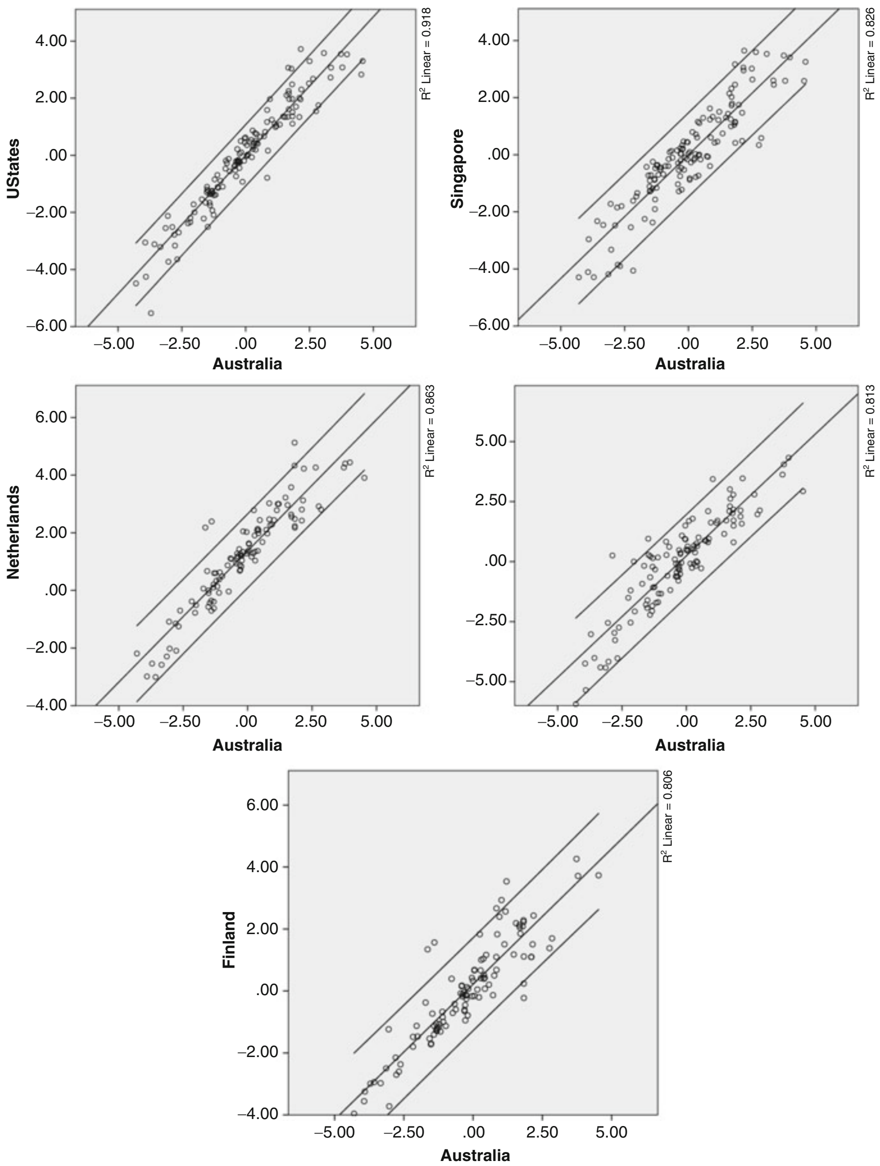


Fig. 7.12 Scatterplot of item delta differences between other countries and Australia

bands help to identify statistically significant outliers in the scatterplots. Some indicators were only reported for small numbers of students. In some cases fewer than five students demonstrated the indicator within a specific country and this would not be sufficient to obtain an accurate estimate of difficulty for that particular indicator in that country. So these might not be 'real' outliers'. The discussion of the differences between countries will continue in Volume 3 of this series.

Interpreting and Reporting for Student Outcomes

The students were administered sets of tasks. Because no previous data exists on these measures it has not been possible to determine benchmark values for interpretive or standard setting purposes. The policy underpinning the project and its manner of reporting to students and teachers mandated that outcomes were not represented as scores, but as descriptions of the skills, knowledge and competencies that a person, regardless of age or education level, ought to demonstrate in order to be given the opportunity to improve performance. These different aspects of social and cognitive skill development, as well as the way in which the levels of readiness to learn have been arrived at, became the core substance of the reporting modules developed as part of the ATC21S project. Further explanations of the reports are provided in Woods, Mountain and Griffin (2015; Chap. 14). The curriculum implications of these levels, as well as the learning readiness levels in general, feature as important aspects of learning and teaching in 21st century schools.

Developing a Skills Progression

The indicators were ordered based on difficulty parameter estimates. The hierarchy of the descriptors of indicators, ordered according to relative difficulty, was again assessed to establish if they were interpreted in a cohesive manner within the broader CPS framework. An iterative process ensured that the conceptual descriptors were supported by indicator location, which in turn used the descriptions of the indicative behaviours to inform the interpretation of the construct continuum. In the same process, the review clarified which indicators had parameter estimates (or locations on the variable) that did not accord with the theoretical model. Given that these were not informative or meaningful within the overall structure of the construct, they were deleted.

Once ordered, the remaining indicators were sub divided into their two dimensions – social or cognitive. Next, levels of the construct [or the developmental progression] were interpreted to help teachers identify relevant groups of students ready for targeted instruction at approximately the same position on the variable. Within both dimensions five levels could be identified to represent the progression from '*novice to expert*'. Indicators which were shown to have little value or influence on the interpretation were also removed. The project team aimed to identify approximately 20 descriptors or performance indicators within each level of the developmental progression.

The data were also interpreted as a single dimension and this identified five levels of skills progression or stages of increasing competence.

According to (Eisner 1993), the use of criteria involves an exercise of judgement, being able to give reasons for the judgment and having an expert understanding of the relevance of criteria to the area being assessed. A teacher must be able to make a judgment about whether a student is at a high or low level of competence, be able to explain the decision, and have an expert knowledge of the domain where these distinctions are made. Similarly, when defining levels on a latent variable (construct or developmental progression) the teacher needs to have the substantive expertise to define the levels and to be able to defend the thresholds between them. In other words teachers will need to develop expertise in collaborative problem solving to use the information provided effectively.

This requirement also demands much more of the assessment task developer than a count of the indicators demonstrated and a declaration of competence according to a pre-defined percentage score which might be established as a pass mark or desirable score or standard. If the data is to be used to describe the level of collaborative problem solving competence, they must enable the teacher to define the range of levels and the criteria or thresholds that separate the levels on the continuum, or the stages of increasing competence. Moreover, it is important to be able to defend the decision to place a student at one of those levels, in terms of what the student can do, not in terms of the number of indicators demonstrated correct. This places serious demands on the measurement expertise of the developer especially when there is no requirement for the students all to demonstrate performance on the same sets of indicators. There must be a theory of development, an understanding of the kinds of tasks that are indicative of progress and a capacity to observe the students' performance on those indicators. These observations also have to occur in circumstances that enable an inference of competence to be made beyond the sample of CPS indicators used. The inference made is the basis of the validity (Messick 1994). Developers or alliances which ignore these requirements and use simplistic coding systems may be providing false information to users and systems of education.

These are demanding requirements for test developers, but are reasonable. They can be supported by a set of assumptions that can combine both criterion referencing and item response modelling to help define a variable or construct and the levels or stages of increasing competence.

Assumptions

- A set of underlying continua can be constructed that describe development or growth in specific domains of learning. The continua define constructs that are measurable, and have direction and units of magnitude.
- The continua do not exist in and of themselves, but are empirically constructed to assist in explaining observations of learned behaviour.

- Each continuum can be defined by a cohesive set of indicative behaviours representing levels of proficiency in the area of learning. These behaviours can be demonstrated through the performance of representative tasks and can be regarded as either direct or indirect indicators of proficiency.
- Not all behaviours can be directly observed. Related, indirect behaviours can be used, along with directly observable behaviours, to describe competency or ability.
- The indicators (behaviours or task descriptions) may be ordered along a continuum according to the amount of the proficiency, competence or ability required for a satisfactory performance or success on each task.
- People can be ordered along the continuum according to the behaviours they are able to exhibit or the tasks that they are able to perform. The behaviours, in turn, can be interpreted to provide a substantive interpretation of the level of proficiency or ability.
- It is not necessary to identify or to observe all possible behaviours or indicators in order to define the continuum. The continuum can be defined by any representative, cohesive sample of indicators that covers a range of levels on the continuum.
- There is no one correct sample of indicators, tasks, test items or behaviours that exclusively defines the continuum or the domain, although there may be a set of indicators that is generally agreed upon as important in defining the continuum.
- While the indicators used to define the continuum are related, there is no causal or dependent relationship between them. It is neither necessary nor obligatory to observe lower order indicators in order to observe higher order behaviours. The existence of higher order indicators implies the ability to demonstrate lower order indicative behaviour. The relationship is probabilistic, not causal.

Standards and Benchmarks

The words "*stages along progressions of increasing competence*" (Glaser 1963, 1981) are important in assessment design and calibration. Criterion referenced or standards referenced interpretation provides an opportunity to link the position of a person or an indicator on a variable (as shown in the variable maps) to an interpretation of what a student, or groups of students, can do, rather than focusing on a score or performance relative to a percentage or a group. The procedure gives substantive interpretation of the levels of increasing competence.

No discussion of developmental growth or progression through stages of increasing competence is complete without a mention of standards. Standards for today will not be acceptable in ten years' time. In the context of 21st century skills rapid change in expected levels of competence is anticipated as the skills become more familiar in education and the workplace. The threshold used today to make a distinction between 'adequate' and 'unacceptable', or between 'mastery and 'non mastery' is established using the experience of a group and an understanding of what can be reasonably expected. It is essentially a norming process, based on an understanding of the tasks and the expected developmental rates of cohort of students. Standards are set to help establish the idea that there is a minimal level of performance that is acceptable by individuals, workplaces and education systems. A standard is one threshold on a continuum chosen to represent some specific meaningful level of achievement. The threshold used to make a distinction between 'adequate' and 'unacceptable' or between 'competent' and 'not competent' is sometimes established using the experience of an expert group and an understanding of what reasonably can be expected. A standard is therefore one threshold in a criterion-referenced framework, which is used to make decisions at a particular point in time. In the context of 21st century skills this may be impossible given the rapid change of the effect and impact of technology on education and workplace skills.

Standards are commonly adjusted as better or higher quality outcomes are desired. The very fact that their improvement is acknowledged and expected means that there is an implicit acknowledgement of the existence of the underlying continuum, which allows the standard and hence ability levels to 'move'. When it is clearer how expectations of 21st century skill levels change, it will be easier to preempt these changes by developing better teaching and learning programs and better ways of setting standards.

Cognitive and Social Competence Levels

The collaborative problem solving skills progressions have been described overall as one dimension in Table 7.10. The second model presented examines both the social and cognitive dimensions separately (Table 7.11). Finally a five dimensional skills progression is described (Table 7.12), with levels relating to the five strands described previously (Table 7.1). These substantive interpretations were produced for cognitive and social skills from the CPS tasks and for strands of perspective taking, participation, social and task regulation, and knowledge building. Performance descriptions were referenced to specific levels that would enable stages of increasing competence on developmental continua to be identified, interpreted and used for instructional intervention.

A primary purpose of the data analysis within the ATC21S project was the identification of skill levels (or levels of competence) in cognitive and social developmental continua displayed by the students. Each of these levels of competence was identified and interpreted by expert panels. The level reached by a person indicates the level of competence that the person typically demonstrates in completing the CPS tasks. Given the method used in the ATC21S project, it also indicates the instructional intervention point for teachers. This information has consequences for teaching intervention and curriculum planning at the school level as well as for the national curriculum – when distributions of students across these levels is identified from proper probability samples. At the school level, for instance, teachers need to focus instruction on the level at which the student is placed, consolidate the level

Table 7.10 One dimension interpretation of collaborative problem solving

Level	Level title	Level description
6	Strategic approach to problem via a collaborative process	The student works collaboratively through the problem solving process and assumes group responsibility for the success of complex tasks. The student works through the problem efficiently and systematically using only relevant resources. They tailor communication, incorporate feedback from their partner and resolve conflicts.
5	Efficient working partnership	The student's actions appear planned and purposeful, identifying cause and effect and basing their goals on prior knowledge. The student promotes interactions and responds to their partner's contributions but may not resolve differences. They adapt original hypotheses and uses suitable strategies to gain a correct path solution for more complex tasks.
4	Cooperative planning	The student perseveres, through multiple strategies, to successfully complete subtasks and simpler tasks. They have developed awareness of their own and their partner's performance abilities. They strive to achieve common understanding and increase co-working by planning strategies and refining goals with their partner. The student adopts a sequential approach and can identify connections and patterns between multiple pieces of information.
3	Awareness of partner & directed effort	The student recognises their partner's significant role in solving the problem and demonstrates effort towards solving the problem. They realise they do not have all the required information and begin to share resources and information with their partner, but with no regard for relevance. They report their own activities and make contributions for their partner's understanding.
2	Investigating the problem	The student attempts to better understand the problem through limited analysis. They assess and utilise their own resources, begins testing hypotheses, and generating broad goals. Interaction with their partner is limited to significant events.
1	Independent inefficient exploration	The student explores the problem space independently with no evidence of collaboration. Their approach is unsystematic and focusing on isolated pieces of information. Interaction with their partner is limited to brief acknowledgements.

below, and set the goal to reach the next level in the continuum. This is what Vygotsky meant by scaffolding. At a national level, new resources (teaching/learning materials) need to be produced, and teacher training reviewed in order that teachers may deal with the different levels appropriately. The ATC21S system includes a series of professional development modules for this reason and these are explained in detail in Woods et al. (2015).

It can be seen from the description of the five dimensions that there are no activity or behavioural indicators of participation at the top level of the participation developmental progression. There is an apparent ceiling of participation skills at the fifth level of the progression. Recall that when indicators were calibrated using a five dimensional, one parameter item response model, the reliability of person

Table 7.11 Two dimensions – social and cognitive, interpretation of collaborative problem solving

Level	Level title	Social	Level title	Cognitive
6	Cooperation and Shared Goals	At this level, the student works collaboratively through the problem solving process and assumes group responsibility for the success of the task. Feedback from their partner is incorporated and used to identify solution paths or modify incorrect ones. The student can evaluate their own and their partners performance and understanding of the task. The student may tailor their communication and manage conflicts with partner successfully, resolving differences before proceeding on a possible solution path.	Refined Strategic Application and Problem Solving	The student's sequential investigations and systematic behaviour require fewer attempts for success and are completed in an optimal amount of time. The student works with their partner to identify and use only relevant and useful resources. The student has a good understanding of the problem and can reconstruct and/or reorganise the problem in an attempt to find alternative solution paths.
5	Appreciated and Valued Partnership	At this level, the student is able to actively participate in scaffolded and unscaffolded environments. The student initiates and promotes interaction with their partner and acknowledges and responds to contributions from their partner. Despite efforts, differences in understanding may not be fully resolved. The student is able to comment on their partner's performance during the task.	Efficient Working	At this level the student's actions appear to be well thought out, planned and purposeful, identifying the necessary sequence of subtasks. The student identifies cause and effect, basing their goals on prior knowledge and uses suitable strategies to gain a correct path solution for both simple and complex tasks. The student can modify and adapt their original hypotheses, in light of new information, testing alternatives hypotheses and adapt additional or alternative of thinking.

(continued)

Table 7.11 (continued)

Level	Level title	Social	Level title	Cognitive
4	Mutual Commitment	At this level, the student perseveres to solve the task as shown by repeated attempts and/or multiple strategies. They share resources and information with their partner and modify communication where necessary to improve mutual and common understanding. Students have an awareness of their partner's performance on the task and can comment on their own performance.	Strategic Planning and Executing	At this level the student can identify connections and patterns between multiple pieces of information. The student is able to simplify the problem, narrow their goal focus and increase co-working by planning strategies with their partner. The student adopts strategic sequential trials and increasing systematic exploration. The student can successfully complete subtasks and simpler tasks.
3	Awareness of Partnership	At this level, the student demonstrates effort towards solving the problem. They become aware of their partner's role in the collaborative problem solving process and recognise the need to engage with their partner. They discuss the task with their partner and make contributions to their partners understanding. The student reports to their partner regarding their own activities on the task.	Sharing and Connecting Information	At this level the student recognises the need for more information, realising that they may not have all the required resources and allocates their own resources to their partner. They attempt to gather as much as possible and begins connecting pieces of information together.

(continued)

Table 7.11 (continued)

Level	Level title	Social	Level title	Cognitive
2	Supported Working	The student actively participates in the task when it is scaffolded but works largely independently. Communication between partners occurs more frequently but is limited to significant events and information necessary to commence the task.	Establishing Information	At this level, the student identifies possible cause and effect of actions, demonstrates an initial understanding of the task concept and begins testing hypotheses and rules. The student limits their analysis of the problem, using only resources and information they have. The student also remains limited in their goal setting generating broad goals.
1	Independent Working	At this level, the student commences the task independently with limited interaction from partner, mainly prompted by instructions. They may acknowledge communication cues by their partner but have not started to work collaboratively. Most communication occurs at the beginning of tasks and only in those tasks where the instructions are clear.	Exploration	At this level, the student explores the problem space but this is limited to following instructions, adopting a singular approach, and focusing on isolated pieces of information. Trial and error appears random and there is little evidence of understanding the consequences of actions resulting in a lack of progress through the task.

Table 7.12 Five 5 Dimension interpretation of collaborative problem solving

	Social			Cognitive	
Level	Participation	Perspective taking	Social regulation	Task regulation	Knowledge building
6		The student can tailor communication with their partner based on their awareness of their partner's understanding, and are easily understood from the start of the task. The student incorporates the contributions from their partner to suggest new solution paths or modify incorrect ones. They use solutions provided by their partner and work more collaboratively through the problem solving process.	The student assumes group responsibility for the success of the task. They can manage conflicts with partner successfully, resolving differences before proceeding on a possible solution path. The student is able to evaluate their overall performance in the task. They are also able to evaluate the strengths and/or weaknesses of their partner based on their performance during task.	The student's approach to the task is systematic. They continue to engage in sequential investigations and systematic behaviour in subsequent task pages that have increased in difficulty. Their actions require few attempts and are hence completed in an optimal amount of time. They have developed an awareness of which resources are most useful and works with their partner to identify the relevant resources and disregard those that posed no benefit in previous trials.	The student has a good understanding of the problem and can reconstruct and/or reorganise the problem in an attempt to find a new solution path.

5	The student is able to actively participate in scaffolded and unscaffolded environments. The student initiates and promotes interaction with their partner often before entering their own answer.	The student acknowledges and responds to contributions from their partner but does not make changes to their original course of action.	The student attempts to resolve differences in understanding with their partner but resolution of differences are not reached. The student is able to comment on their partner's performance during the task.	The student can identify the necessary sequence of subtasks in order to achieve task completion. Actions appear 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 which may be useful in future tasks/subtasks or for an alternative solution path.	The student can identify cause and effect and use suitable strategies to gain a correct path solution for both simple and complex tasks.
					The student can modify and adapt their original hypotheses, in light of new information, testing alternatives hypotheses and altering their course of thinking.

(continued)

Table 7.12 (continued)

Level	Social			Cognitive	
	Participation	Perspective taking	Social regulation	Task regulation	Knowledge building
4	The student perseveres to solve the task as shown by repeated attempts and/or multiple strategies.	The student modifies communication with their partner to improve mutual understanding and share resources and information.	The student reports to their partner regarding their own activities on the task.	The student adopts strategic sequential trials and increasing systematic exploration. They narrow their goal setting and focus on successfully complete a subtask before moving on. The student simplifies the problem, analysing it in stages and plans strategies with their partner.	The student can identify connections and patterns between multiple pieces of information. The student can successfully complete subtasks and simpler tasks.
3	The student demonstrates effort towards work towards solving the problem. The student discusses the task with their partner by responding to communication cues and requests.	The student makes contributions to their partners understanding.	The student comments on or share information to their partner regarding their own performance while attempting the task. The student is aware of their partner's performance on the task. They can reach a common understanding with their partner in regards to the task.	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.	The student begins to connect pieces of information together.

2	The student is aware of their partner's role. and actively participates in the task when it is scaffolded. Communication between partners occurs more frequently but is limited to significant events and information necessary to commence the task.	The student is not overtly responsive to their partner, often taking a long time to respond or not at all and tends to ignore their partners contributions.	The student still works largely independently taking responsibility for their own actions during the task. The student is aware of their own level of performance during the task.	The student limits their analysis of the problem by only using the resources and information they have and, following system instructions. 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.	The student tests their hypotheses based on the information they have. They identify possible cause and effect of actions and repeats attempts in order to gain more information about an actions outcome.

(continued)

Table 7.12 (continued)

Level	Social			Cognitive	
	Participation	Perspective taking	Social regulation	Task regulation	Knowledge building
1	The student commences the task independently and task exploration is mainly directed by system instructions. The student shows limited interaction with partner. They may acknowledge communication cues by their partner but have not started to work collaboratively (i.e. sharing information or resources). Most communication occurs at the beginning of tasks and only in those tasks where the instructions are clear.			The student explores the problem space by clicking on various resources often in a random fashion. However, if the student has difficulty understanding the task they make very little attempt to explore the problem space. They engage in singular approaches to trial and error in an attempt to build knowledge of the problem space. They attempt to solve the problem through an apparent unsystematic guessing approach and tend to repeat errors or reproduce unproductive actions with no clear indication of advancing through the task within several attempts.	The student continually attempts the task with the same approach with little evidence of understanding the consequences of actions taken. The student focuses on each piece of information individually; only following the specific instructions provided.

separation on the participation strand was lower than that of the other dimensions. This is to be expected if there are no indicators at the top level of the progression, especially considering the students scored highest on this strand. In the context of the task design, the most obvious explanation is that students were strongly motivated to participate by the format of the tasks, but the indicators did not adequately address higher order participation, for example perseverance in extremely difficult circumstances. This is a limitation in the design of this assessment; although it may be a necessary limitation given the other constraints to task design.

At the lower level of both perspective taking and social regulation skills there appears to be no data or indicative actions or behaviours that we were able to identify to define a lower level skill in these progressions. What this means is that perspective taking and social regulation may demand a higher level of ability in collaborative problem solving overall before evidence of these skills can be observed. In other words, some subskills or elements from the collaborative problem solving framework may act as enablers for the development of other subskills.

Level 2 in perspective taking is described as “The student is not overtly responsive to their partner, often taking a long time to respond or not at all and tends to ignore their partners contributions”. Below this level a student would perhaps be completely unaware of their partner. This would assumedly be the lowest level of perspective taking – either ignoring or not understanding that there is a partner. The interactive nature of these tasks limits the ability to measure such a low level of perspective in students. Students are told that there is a partner and the tasks are designed such that the perspective of the partner should be considered. Therefore in terms of perspective it is almost automatic that students must begin on level 2.

In terms of social regulation, it is also not surprising that level 1 is not able to be measured. If we were to measure the lowest level of social regulation, it would also most likely be an absence of understanding or acknowledgment that there was a partner with whom to regulate or to whom to report. Again this could be seen as a limitation of the task design, but may be a necessary limitation given the strands to be measured and the context in which we are measuring collaborative problem solving skills.

These skill progressions have been well developed conceptually, and the descriptors supported by empirical item locations. There may be some limitations as described above, but these are due to the intended design of the tasks. These progressions provide a useful framework for teachers to use in interpreting their observations of student behaviour.

References

Adams, R. J., & Khoo, S. T. (1993). *Quest: The interactive test analysis system*. Melbourne: Australian Council for Educational Research.

Adams, R., Vista, A., Scoular, C., Awwal, N., Griffin, P., & Care, E. (2015). Automatic coding procedures for collaborative problem solving. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills: Methods and approach* (pp. 115–132). Dordrecht: Springer.

Adamson, F., & Darling-Hammond, L. (2015). Policy pathways for twenty-first century skills. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills: Methods and approach* (pp. 293–310). Dordrecht: Springer.

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

Eisner, E. (1993). Why standards may not improve schools. *Educational Leadership, 50*(5), 22–23.

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

Glaser, R. (1981). The future of testing: A research agenda for cognitive psychology and psychometrics. *American Psychologist, 36*, 923–936.

Griffin, P., Care, E., Bui, M., & Zoanetti, N. (2013). Development of the assessment design and delivery of collaborative problem solving in the Assessment and Teaching of 21st Century Skills Project. In E. McKay (Ed.), *ePedagogy in online learning: New developments in web mediated human computer interaction*. Hershey: IGI Global.

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* (pp. 37–56). Dordrecht: Springer.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher, 23*(2), 13–23.

Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests* (Copenhagen: Danish Institute for Educational Research), expanded edition (1980) with foreword and afterword by B. D. Wright. Chicago: The University of Chicago Press.

Wilson, M., & Adams, R. J. (1995). Rasch models for item bundles. *Psychometrika, 60*(2), 181–198.

Woods, K., Mountain, R., & Griffin, P. (2015). Linking developmental progressions to teaching. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills: Methods and approach* (pp. 267–292). Dordrecht: Springer.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis* (p. 1982). Chicago: Mesa Press.

Part IV
Participating Countries in the Field Work

The six chapters from participating countries provide information about how each country implemented the research and development phases within their education systems. The chapters follow a similar structure, with some variations reflecting their particular concerns or activities. The educational context of each country is outlined with particular attention focused on moves within the country toward curriculum or system changes recognising the importance of 21st century skills. This is followed by description of the how the ATC21S implementation was devolved to government departments or research institutions. The four processes of task concept check, cognitive laboratory, pilot and trial are covered to differing degrees by each country, and are followed by a discussion of the challenges experienced by each and, finally, some concluding statements.

Care et al. (2015, Chap. 8), writing on the project in Australia, describe the processes in some detail, with a focus on the guiding questions and data collection materials used to obtain feedback about the approach and the tasks. The description of the processes complements the information provided in Griffin and Care (2015, Chap. 1), and expands on the methods used by all countries across the four phases. Care et al. draw attention to the challenges for ATC21S in Australia presented by the federal versus state responsibilities for education, variations in ICT infrastructures across states and individual schools, and in ICT teacher competencies, and the need for teachers to become familiar with 21st century skills prior to teaching them.

Poon et al. (2015, Chap. 9) outline Singapore's approach to ATC21S participation, using the project as a data source to better understand how teachers and students think about 21st century skills in the educational context. Poon et al. provide details of numbers of schools, teachers and students engaged in the project and illustrate the experiences and views of the teachers and students with selected comments. Within their Method section, responses to the ATC21S tasks are organised across six main themes: relevance of 21st century skills to education: engagement with the tasks, implications of content-rich versus content-free tasks, collaboration capacities, implications of ambiguity in assessment, and tools and technical issues. The chapter concludes with discussion of needs in future development of the innovative approaches demonstrated by ATC21S.

Ahonen and Kankaanranta (2015, Chap. 10), writing on Finland, describe localization and translation processes that were required for their participation in ATC21S. They also focus on the responses of teachers and students to the four phases of the research and development implementation, providing insights into the teacher and student experience. They reflect on the need for a strong research presence in participating schools and the intensive nature of such an activity both in terms of data collection and in terms of contribution of participants' time. The Finnish experience reflects the benefits of continuing involvement of schools and their personnel throughout the process, and how such involvement not only builds engagement but also informs the outcomes of such studies.

Comfort (2015, Chap. 11) contextualizes ATC21S within the framework of other U.S. 21st century initiatives such as Partnerships 21. She then provides grounded detail of the research and development phase with particular attention to teacher and student responses to the Learning through Digital Networks assessment tasks. Her overview provides descriptions of data gathering tools and populates these to illustrate typical responses from participants to the tasks. Comfort concludes with a series of recommendations for the articulation of the process from assessment to teaching of 21st century skills.

Bujanda and Campos (2015, Chap. 12) contextualise the participation and activity of Costa Rica as a lead country in Latin America for the ATC21S approach. Consistent with their engagement, they brought to their participation the additional innovation of re-skinning the tasks which were designed to assess Learning through Digital Networks – Information Communications Technology. The efforts of the Costa Rican team, combining input from the Ministry of Education, the Omar Dengo foundation, and experts from substantive discipline areas provide an excellent illustration of the capacity of this assessment approach to be adapted to different language and cultural environments.

Schonau (2015, Chap. 13) provides an account of the Dutch experience of ATC21S, with its focus on the trial of the collaborative problem solving assessment tasks. Schonau points out some interesting aspects concerning translation of interactive and asymmetric tasks, focusing on the need for equivalent rather than literal translation. These translation issues apply to the actual navigation around the tasks as well as to content that is specific to the tasks themselves. Suggestions are made concerning how to implement research trials in schools in terms of providing additional familiarisation and feedback.

Each of these chapters reflects an individual experience of the common requirements of the global project. Each also varies according to factors of importance to the different countries, such as translation issues for Finland, Costa Rica and the Netherlands; state and federal implications for Australia and the U.S.A; and recommendations for either assessment, as in the case of Singapore, or teaching, as in the case of the U.S.A. The degree to which their teachers' and students' experiences of the tasks through their development are similar is very strong. The implementation of the activities beyond the core global requirements to optional additional training and debriefing provides the opportunity to analyse the degree to which such activities have enhanced the individual country experiences and the consequent value of the project to them.

Care, E., Scoular, C., & Bui, M. (2015). Australia in the context of the ATC21S project. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills: Methods and approach* (pp. 183–197). Dordrecht: Springer.

Poon, C. L., Tan, S., Cheah, H. M., Lim, P. Y., & Ng, H. L. (2015). Student and teacher responses to collaborative problem solving and learning through digital networks in Singapore. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills: Methods and approach* (pp. 199–212). Dordrecht: Springer.

Ahonen, A. K., & Kankaanranta, M. (2015). Introducing assessment tools for 21st century skills in Finland. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills: Methods and approach* (pp. 213–225). Dordrecht: Springer.

Comfort, K. (2015). Case study on the implementation of ATC21S in the United States. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills: Methods and approach* (pp. 227–244). Dordrecht: Springer.

Bujanda, M. E., & Campos, E. (2015). The adaptation and contextualization of ATC21S™ by Costa Rica. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills: Methods and approach* (pp. 245–256). Dordrecht: Springer.

Schönau, D. (2015). ATC21S trials of collaborative problem solving tasks in the Netherlands. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills: Methods and approach* (pp. 257–263). Dordrecht: Springer.

Chapter 8
Australia in the Context of the ATC21S Project

Esther Care, Claire Scoular, and Myvan Bui

> *In the 21st Century Australia's capacity to provide a high quality of life for all will depend on the ability to compete in the global economy on knowledge and innovation.*
>
> *(Ministerial Council on Education Employment Training and* Youth Affairs [MCEETYA] 2008, p. 4)

Abstract Schooling in Australia involves 1 year in preparatory school and 12 years of primary and secondary school, with compulsory education to 15 years of age (Australian Government, Department of Education, Employment and Workplace Relations. http://www.deewr.gov.au. Accessed 8 Nov 2012, 2012). Education can be undertaken in government or non-government (independent or Catholic) schools. The Australian Constitution allocates responsibility for the operation of schooling to the state and territory governments (Australian Government, Review of funding for schooling. Emerging issues paper. Commonwealth of Australia, Canberra. http://www.deewr.gov.au, 2010b). The Australian Government (Review of funding for schooling. Emerging issues paper. Commonwealth of Australia, Canberra. http://www.deewr.gov.au, 2010b) is responsible for providing national leadership in educational reforms, and investing substantial funding in their delivery. An example of such leadership was the development of a national curriculum for which the states and territories bear responsibility for implementation. All state and territory Ministers with responsibility for education, and the Federal Minister for Education, are brought together in a Ministerial Council approximately every 10 years to determine strategies for the future of Australia's education systems. The strategies are published as a 'declaration' after each Ministerial Council, and the 2008 *Melbourne Declaration on Educational Goals for Young Australians* (MCEETYA, Melbourne declaration on

E. Care (✉) • C. Scoular • M. Bui
Assessment Research Centre, Melbourne Graduate School of Education,
University of Melbourne, Parkville, VIC, Australia
e-mail: e.care@unimelb.edu.au

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_8

educational goals for young Australians. MCEETYA, Melbourne. Retrieved from http://www.mceecdya.edu.au/verve/_resources/National_Declaration_on_the_Educational_Goals_for_Young_Australians.pdf, 2008) initiated the redesign of the curriculum to develop successful learners, who are expected to have the skills necessary to be "creative and productive users of technology, especially ICT, as a foundation for success in all learning areas" (MCEETYA, Melbourne declaration on educational goals for young Australians. MCEETYA, Melbourne. Retrieved from http://www.mceecdya.edu.au/verve/_resources/National_Declaration_on_the_Educational_Goals_for_Young_Australians.pdf, 2008, p. 8). Under the National Curriculum, the general capabilities to be taught across all learning areas include ICT competence and the ability to work collaboratively in teams, across cultures and disciplines.

Context

Between 2009 and 2012, the Australian Government budgeted to invest $65 billion in Australian schools (2012). This funding was divided across a number of initiatives including the Digital Education Revolution (DER; Australian Government 2012).

The DER was a 5-year federal government election commitment made in 2008. Its primary goal was to enable schools to gain access to information technology for students (Australian Government 2008). The DER policy commitment made explicit that computer equipment would be provided to schools through an allocated funding stream and that upgrades to the national infrastructure would be supported (Australian Government 2008). Between 2010 and 2012, the Australian Government (2010a) committed funding to meet a student-to-computer ratio of 1:1 by the end of 2011, and teachers and school leaders were also provided with support to implement and integrate ICT in the classroom. Implementation of the DER required collaboration across education jurisdictions to ensure effective sharing of the resources, tools and expertise (Australian Government 2008). The DER encompassed four areas of change: leadership, infrastructure, learning resources, and teacher capability.

One of the major requirements for the DER to be successfully implemented is the provision of a nationwide high-speed broadband network. The National Broadband Network (NBN) is intended to meet this requirement by delivering high-speed broadband connectivity of the type that will be necessary to support the ICT tools needed for education in digital technology (Australian Government 2011). Surveys were conducted over 3 years, from 2008 to 2010, to identify the connectivity needs in schools. Each survey collected information about the technology types used in schools (i.e., fibre, copper, satellite, wi-fi, etc.), the bandwidth available and the service providers used (Australian Government 2011). The results of the 2010 survey indicate that there was an improvement or increase across all three areas in each year of the survey (Australian Government 2011). Programs to meet the teacher capability and leadership goals of the DER were also put in place.

PLANE (Pathways for Learning, Anywhere, anytime – a Network for Educators) was one of the projects funded by the Australian Government to help educators develop skills, knowledge and experience in teaching ICT and encouraging its use by students. PLANE provides an online professional learning environment which seeks to improve teachers' confidence to use technology in the classroom and thereby promote the development of digital literacy and 21st century learning skills. The Teacher Online Toolkit is a component of this, and it seeks to "enhance the capacity of in-service teachers to effectively incorporate varied technologies in classrooms while assisting the implementation of the Australian Curriculum" (Education Services Australia n.d., p. 1). Tools for pre-service teachers are provided through the *Teaching Teachers for the Future* initiative, which focuses on increasing ICT proficiency in graduate teachers across Australia (Australian Government n.d.).

The introduction of a national curriculum in Australia provides a formal framework within which the assessment and teaching of 21st century skills will take place. In addition to traditional discipline, or key, learning areas, such as mathematics, history and science, the national curriculum being phased in from 2012 onward includes a set of seven "general capabilities" which closely approximate what are referred to as 21st century skills or competencies across initiatives such as ATC21S,[1] Partnerships 21, and UNESCO's Delors Report (1996). The Australian team mapped these capabilities back to the ATC21S KSAVE framework, demonstrating the strong alignment between the national initiative and the global ATC21S project.

ATC21S Development in Australia

ATC21S is one of the initiatives undertaken by the Australian Government through the former Department of Education, Employment and Workplace Relations (DEEWR), and specifically its Branch concerned with School Performance and Improvement. The Branch implemented the initiative through two actions. In order to engage Australia-wide participation through the states, they approached the Australian Education, Early Childhood Development and Youth Affairs Senior Officials Committee (AEEYSOC), which supports The Ministerial Council for Education, Early Childhood Development and Youth Affairs, whose responsibilities are: primary and secondary education; youth affairs and youth policy relating to schooling; cross-sectoral matters including transitions and careers; early childhood development, including early childhood education and care; and international school education. Recruitment for participation in the project took place in the first instance through this Committee. The second action was to engage a National Research Coordinator (NRC) to carry out the Australian arm of the project – the Coordinator in this case being based at the University of Melbourne, which was also coordinating the global project. Throughout the project, the NRC reported to and was supported by DEEWR staff, who were actively involved in the project by promoting it among the states and territories, and identifying synergies of ATC21S with major reform in the Australian national curriculum.

---

[1] The acronym ATC21S™ has been globally trademarked. For purposes of simplicity the acronym is presented throughout this chapter as ATC21S.

Method

With its twin focus on assessment and teaching of Learning through Digital Networks – Information and Communication Technology (LDN-ICT) and Collaborative Problem Solving (CPS) skills, the ATC21S project provided the opportunity to develop 21st century skills among school students, and awareness of the importance of ICT and collaborative problem solving skills among teachers and school leaders. The ATC21S Assessment Task System was developed in-house at the Assessment Research Centre, University of Melbourne, in 2010/2011 and this prototype system was used in the research and development process of designing and presenting assessment tasks to Australian students in schools in the states of Victoria, Western Australia, Queensland and NSW. The project's principal objective was to test and validate assessment materials in Australian schools, and it was undertaken in four phases:

1. Task concept check
2. Cognitive laboratories
3. Pilot study
4. Field trials

The National Research Coordinator (NRC) was tasked with organising links with schools, managing data collection activities in schools – including appropriate quality assurance of data – and liaising with the International Research Coordinator (IRC) to deliver the outcomes. The NRC also liaised with the Australian National Project Manager (NPM; DEEWR) to report on progress throughout the project.

A liaison person within each state department of education was appointed to manage recruitment of schools to the project. The liaison person disseminated information about ATC21S and subsequently nominated schools to be contacted by the NRC team. Once communication was established with individual schools, the NRC team worked with school leaders and individual teachers to arrange workshops with teachers and assessment activities with students.

Task Concept Check

The Concept Check phase was designed to ascertain whether the assessment task ideas for LDN-ICT Literacy and Collaborative Problem Solving appeared reasonable to teachers in schools. To do this, 16 teachers of students aged between 11 and 15 years from Victorian schools were asked to indicate whether each task concept looked as though it would:

- Engage the students;
- Be appropriate for the students in terms of prerequisite knowledge;
- Be appropriate for the students in terms of socio-cultural context;
- Take similar or different amounts of time for the students; and
- Differentiate between students.

The teachers were asked not to focus on actually completing the task (nor attending to its detail), but to engage with it enough to determine its conceptual characteristics and the manner in which these might affect a student's collaborative interaction with the task. While responding to questions from the interviewer, participants checked the draft items. Subsequent group discussion with the teachers used the items in Table 8.1 to stimulate reflection on the task concepts.

Table 8.1 Task concept check – focus questions for teachers

1. What skills or capabilities do you think the tasks are targeting?
Participation
Activity within environment
Interacting with, promoting, and responding to the contributions of others
Understanding and completing a task or part of a task individually
Perspective taking
Ignoring, accepting or adapting contributions of others
Awareness of how to adapt behaviour to increase suitability for others
Awareness of how to contextualise the contributions of others
Awareness of how to tailor own contributions to intellectual capabilities of others
Social regulation
Achieving a resolution or reaching a compromise
Recognising own strengths and weaknesses
Recognising strengths and weaknesses of others
Assuming responsibility for ensuring aspects of task are completed by the group
Managing divergence (of opinions, viewpoints, concepts) by negotiating with others
Problem solving
Setting a clear goal for the task
Managing people or resources to complete a task
Formulating a course of action to address a problem or task
Implementing possible solutions to a problem and monitoring progress
Analysing and defining a problem in familiar language
Identifying need for further information
Developing lines of argument and explaining ideas to others
Accepting ambiguous situations and exploring options within these
Changing from one line of reasoning or course of action to another as information or circumstances change
Knowledge building
Making connections between elements of knowledge
Following a path to gain knowledge
2. Considering the capabilities of your students, are there any questions or activities that seem too far above or too far below the general range of capabilities?
3. What other comments do you have about the task?

Cognitive Laboratory

The purpose of the cognitive laboratories ("coglabs") was to check the assessment tasks in their alpha versions and ensure that they had the capacity to elicit evidence of the two skillsets from students, and that this evidence could be mapped on to the two skills frameworks.

In response to the global research requirement, each country was to ensure that for the two skillsets, each assessment task was tried by at least three pairs of students in each age range of 11 years, 13 years and 15 years. Accordingly in Australia, the initial cognitive laboratories involved 62 students aged 11–15 years from schools in Victoria, Queensland and NSW. With support from the NRC team, each student attempted at least one assessment task, with students paired to undertake tasks collaboratively. Altogether, there were 16 cognitive laboratory sessions. Data were captured in three main ways: by the NRC Team observing and making hand-written notes on forms designed for the purpose; through full screen and audio capture via Camtasia software; and from focus group sessions with students following the online task administration. Each session lasted about one and a half hours, including the briefing to students and post-task interview.

In order to obtain the richest information possible from the students, the observers used the following points as a guide during the test administration.

- Sit near the student but not in their personal space
- If the student is silent for more than a few seconds, prompt with "Keep talking" or "What's happening"
- Be sure that the student is actually entering her/his responses and/or taking action, not just talking, in which case prompt with "Please enter your response"; if the student is having trouble entering responses and talking simultaneously, have the student talk first and then enter her/his responses
- If the student asks you what to do because s/he does not understand a question, tell her/him: "Ask your partner or do whatever you think makes sense"; you should not help them solve the problem.
- Be attentive with body language by head-nodding and occasional, non-evaluative comments in response to students
- Do NOT tell the student if s/he is getting an answer right or wrong
- Do NOT tell the student if s/he is doing well/poorly on the activity
- Do NOT show bias for certain tasks, items or item formats (e.g., do not say anything like, "This is not a very good problem" or "Problems like these don't test many skills").

Observers were provided with different lists of behaviours to audit depending on which skillset was being observed – collaborative problem solving or LDN-ICT literacy. For example, for collaborative problem solving, observers took notes to populate a response table which included the behaviours listed in Table 8.2. For the LDN-ICT Literacy tasks, both specific and general behaviours were monitored. Table 8.3 provides an example of actions of particular interest. The observers were required to identify which behaviours occurred for different pairs of students, here

Table 8.2 Cognitive laboratory checklist items for observations of students engaging in collaborative problem solving tasks

Did the student:
Show engagement with [this page of] the task
Interact with their partner without prompting by the observer
Interact with their partner with prompting by the observer
Agree with their partner on a solution to the problem
Complete the tasks on this page
Work on the tasks on this page mainly individually
Work on the tasks on this page mainly with their partner
Take the lead in the tasks on this page
Follow their partner's lead or suggestions
Ignore their partner's lead or suggestions
Suggest a solution to the problem to themselves or their partner
Suggest more than one solution to the problem to themselves or their partner
Show frustration with the task
Show frustration with their partner
Show signs of enjoyment in the task
Comment positively on their partner's behaviour to themselves or their partner
Comment negatively on their partner's behaviour to themselves or their partner
Comment positively on their own behaviour to themselves or their partner
Comment negatively on their own behaviour to themselves or their partner
Take a systematic approach to solving the problem
Give to themselves any reasoning for their actions
Give to their partner any reasoning for their actions
Were there Usability Issues (specify)?

identified as Aus1trek15, etc. For this stage of item/test development, it was important to identify the degree to which student pairs of different abilities interacted with the task. In addition, this monitoring provided another check of the recording of actions in the logfiles against observed performance.

After the assessments, researchers conducted short focus group sessions with the students. Questions included seeking information about whether students thought the tasks would provide "fair" assessments of their skills, what skills the students thought were being assessed, whether the tasks were enjoyable, what they thought might have contributed to their progress, and whether they had learnt anything from the experience. In addition, based on observer judgment, problematic questions or issues for students were reviewed with the students by going back into the task environment to discuss points at which observers did not understand what the student was doing or at screens where the student/s had seemed to struggle. Prompts

Table 8.3 Part of cognitive laboratory observation procedure form for one page of Arctic Trek, an LDN-ICT Literacy task (cognitive laboratory version)

Task ID: 69	Clue 3 overview		☐ Use slider to indicate polar bear population. ☐ On first box, post individual answer ☐ Cut and paste answer from team notebook and enter on second box					
	Student login IDs							
Actions taken	**Aus1trek15**		**Aus2trek15**		**Aus3trek15**		**Aus4trek15**	
Click relevant link	Y☐	N☐	Y☐	N☐	Y☐	N☐	Y☐	N☐
Identify and count colours	Y☐	N☐	Y☐	N☐	Y☐	N☐	Y☐	N☐
Use slider to select a number	Y☐	N☐	Y☐	N☐	Y☐	N☐	Y☐	N☐
Access Google Docs Notebook	Y☐	N☐	Y☐	N☐	Y☐	N☐	Y☐	N☐
Enter text in upper textbox	Y☐	N☐	Y☐	N☐	Y☐	N☐	Y☐	N☐
Post answer on team notebook	Y☐	N☐	Y☐	N☐	Y☐	N☐	Y☐	N☐
Copy/paste text in textbox	Y☐	N☐	Y☐	N☐	Y☐	N☐	Y☐	N☐
Click Get Hint if applicable	Y☐	N☐	Y☐	N☐	Y☐	N☐	Y☐	N☐
Teach aid	Y☐	N☐	Y☐	N☐	Y☐	N☐	Y☐	N☐
Usability issues	Y☐	N☐	Y☐	N☐	Y☐	N☐	Y☐	N☐
Comments								
Login ID: Aus1trek15								
Login ID: Aus2trek15								
Login ID: Aus3trek15								
Login ID: Aus4trek15								

in this process included "What do you think this question is asking you to do?" and "How could we make any wording of the task clearer?" Students were also asked whether they had taken similar online tests previously, and about technical and technological difficulties or challenges they had encountered.

Responses from the Task Concept Checks and Cognitive Laboratories

Feedback was summarised and then integrated with responses from the other countries participating in the global project. The synthesised comments were provided to the task development agencies for consideration in their building of the interactive assessment tasks.

Two major issues were identified through these processes. The first revolved around clarity of the tasks. Students are typically accustomed to having assessment tasks described to them very clearly so that they know what they are expected to do. The tasks developed for these measures did not meet this expectation, which confused the students. They lacked initial understanding about how to proceed with the tasks, did not realise that exploration was a required activity in order to understand the problem space, and assumed that working towards a correct answer was a primary concern. In terms of understanding the problem space in the case of collaborative problem solving, the students did not initially understand that the partners in each pair were seeing different to images. Similarly they did not invariably understand that it was both acceptable and desirable that they help each other. Teachers felt that insufficient information about the task would lead students to wonder why they were doing it, and hence to lose focus or motivation. It was acknowledged that a degree of ambiguity was useful from a problem solving point of view but teachers thought that students would need more guidance as to the purpose and goals of the activity, as well as reassurance in the form of rewards or feedback as they progressed through the activity. A common misconception among teachers was that students would be 'graded' on their performance and that they should therefore be given every opportunity to excel at the tasks. In reality, the system was designed to measure how well students were able to seek solutions to problems in a collaborative manner when given imprecise and incomplete information. For this reason, suggestions that the task instructions be made clearer, or that more guidance be provided, were not necessarily incorporated in the online task refinement. However, a guidance manual was written for teachers to understand the nature of the tasks. It outlined the suggested advice to provide to students commencing the tasks in order to ensure that the students were comfortable with the nature of the task but not provided with too much scaffolding.

The second issue revolved around technical issues and connectivity. Some technical issues arose from the lack of clarity discussed above – because students expected to be told exactly what to do, they did not explore the space and therefore did not understand the flexibility of action that was required. This had impact when students were unsure how to move from one part of a task to another, or from one page to another. Early versions of the tasks also had some bugs and delays in connectivity, which exacerbated some student frustration.

Some feedback from teachers suggested that certain tasks appeared too mathematical or science-based, and teachers who did not teach these subjects were concerned that the tasks were outside their area of expertise. They suggested that students might apply a maths subject approach to tasks that involved the manipulation of numbers, and this would affect their attitude to the task. They noted that students would typically assume when interacting with a mathematical task that there was a right or wrong answer. When incorporating feedback into the design of the system, teachers' views were taken into account in the context of a broader understanding of the purpose of the tasks. The tasks were subsequently developed so that content-based tasks did not require a knowledge-based approach, even though a basic content understanding might prove advantageous. These findings were very useful for task development but also highlighted that the very nature of the skills being assessed did require that tasks not be fully scaffolded for the students, since this would nullify their usefulness and the validity of assessment results.

Pilot

The pilot phase enabled researchers to record teachers' ease with the process of implementing online assessment sessions in classrooms, and to test the collection of whole class student data. The pilot was conducted across several months in line with the development of assessment tasks for the two skill sets. Seven Australian schools – approximately 108 students – participated in the pilot. Grade level selection varied across schools in response to school preferences, school characteristics and logistic imperatives. The students were spread across ability levels, given that non-select intake schools participated in the pilots, and full non-streamed classes took part. The process was designed to fulfill two main functions:

1. Logistics – identify that the tasks can run seamlessly in a mainstream classroom context and that teachers are able to administer them;
2. Pilot data – collect data to enable the development of draft empirical learning progressions to guide the process of the development of scoring algorithms.

Prior to the pilot sessions, 30–45 min informal preparation sessions were conducted at the schools during which the researcher outlined the process of administering the tasks. The teachers were also provided with an administration manual. The manual provided an introduction to the two skill sets and the assessment tasks; information about technology requirements in terms of hardware, software and internet access; the process for allocation of student logins and registration; and the actual classroom administration procedures, including sequence of tasks and timelines.

Researchers observed the pilot sessions being conducted and recorded teachers' ease with the process, as well as information about student experience and responses. Questions to which answers were required are listed in Table 8.4.

Trials

The fourth phase, field trials, was conducted in two parts. In the first part, conducted in 2011, trials of the Flash version of the collaborative problem solving tasks were held. The purpose of this phase was to inform the scoring of the tasks and the efficacy of the coding system. Large numbers of student participants were required to establish the validity of empirically based scales that have the capacity to locate students on the developmental continua associated with each of the skill sets. Thus, this trial involved 20 schools and a total of 660 students aged between 11 and 15 years: 60 students at age 11 years, 300 at age 13 years and 300 at age 15 years. Students who participated in the trial were spread evenly across ability level. This approach was consistent with the goals of the global research effort, in terms of collecting data from students who might be both less and more skilled, in order to develop and finalise coding criteria for scoring. Task administration took place across two 50–60 min class periods. Table 8.5 provides a breakdown of the numbers of completions for

Table 8.4 IT and classroom management checklist for pilots

1. What were the issues that the IT support staff needed to deal with in order to enable access to the tasks, prior to the session?
2. Were there any events/issues, in terms of hardware, software, or internet access that required IT support staff to intervene during the session? If so, what were these?
How were they were resolved?
How long did this take?
What were the repercussions for management of the class?
What were the repercussions for individual students?
3. How long from the time students were at their desks, did it take for all students to be logged into the task site?
How much time were students able to spend on the actual assessment questions?
4. Were there any difficulties for specific students in engaging with the online environment? If so, what were these?
Were these difficulties associated with individual student characteristics such as physical or cognitive disabilities or dysfunctions?
Were these difficulties associated with technical issues? For example, slow connection speeds, inappropriate screen sizes, error messages?
5. Did the teacher experience any difficulties in terms of communicating to students what would be required of them?
6. Did the teacher experience any difficulties in terms of classroom management during the session?
7. To what extent did students completing tasks at different times prove problematic in the classroom?
8. To what extent was the need to group students for tasks managed effectively?
What strategies were implemented to facilitate the task/student allocation?
9. Is it viable to run such a session with one staff member only in the classroom?
10. What advice, comments or suggestions, if any, did the teacher have for the administration of the session?
11. For future classroom administration of tasks, what are the most important strategies or structures you would put in place to enhance the smooth running of the class?
12. Complete the attached worksheet concerning composition, structure, timing of the class student numbers, tasks completed

each task, along with figures for the second part of the trials, conducted in 2012 and discussed below. Note that each student completed more than one task.

The second part of Phase 4 involved reconfiguring the system to overcome several issues experienced in the Flash format (see Awwal et al. 2015; Chap. 5). To achieve this, the assessment tasks were re-programmed in HTML5, within a registration system that could be managed at different levels – international, within country, within jurisdictions within country, and within schools within jurisdictions. During 2012, only the assessment task component of this system could be programmed for trial. Due to delays in system-readiness, the trials could not commence until late September 2012, which limited opportunities to recruit schools, as they were heading towards the end of the school year. Nevertheless, a total of 13 schools in Victoria and Western Australia participated, and Table 8.5 provides details of the numbers of task completions, broken down by age.

Table 8.5 Task completions by Australian students in 2011 and 2012 trials (note that each student completed more than one task)

	Australia							
	11 year old		13 year old		15 year old		Overall	
Task name	2011	2012	2011	2012	2011	2012	2011 total	2012 total
Olive oil	311	0	105	154	76	56	492	210
Laughing clowns	310	0	125	338	63	96	498	434
Hot chocolate	280	0	105	108	66	36	451	144
Shared garden	17	0	69	182	30	78	116	260
Small pyramids	34	0	81	182	52	50	167	232
Sunflower	257	0	85	178	67	60	409	238
Balance	129	0	92	118	53	48	274	166
Plant growth	195	0	87	186	79	54	361	240
Warehouse	179	0	107	88	79	48	365	136
Hexagons	12	0	89	152	55	50	156	202
Game of 20	33	0	118	136	64	24	215	160
Practice task	307	0	233	0	144	0	660	0

Of participating students in these trials, 38 % were in year 7, 37 % were in year 8, 17 % were in year 9, and 7 % were in year 10. The same recruitment procedure was used as in 2011, but this time the teacher training was conducted over the telephone, and teachers were tasked with running the trials in class. To facilitate this, a step guide to administration was developed, which provided a simple road map for teachers to follow when running the system trials in class. Login pairs and bundles of tasks were provided to the teachers in advance of each session, and feedback was requested from teachers after the trials were completed.

Responses from the Pilot and Trials

The most common piece of feedback from the teachers as a result of the pilot and trials was that they were uncertain regarding the constructs and the skills being measured in the tasks. Teachers were given only a brief introduction to the constructs and the skills being measured, and many reported they felt ill equipped to comment on whether they were able to observe these skills. It was recognised that professional development would be a welcome pre-requisite to running trials in their classrooms. The importance of being able to debrief with students after they have completed the tasks was also recognized. As a result, professional development modules were developed in document format. These included background information on the project, the constructs and skills of LDN-ICT Literacy and CPS, developmental learning, administration of the assessments, interpretation of the reports and strategies for adopting this information in the classroom in an offline capacity.

A major issue affecting the smooth running of the assessment sessions was the use by schools of internet firewalls that can inhibit access to the assessment tasks.

This issue was complicated by the fact that different standards and processes are implemented across different Australian states. The reprogrammed versions of the tasks were most compatible with the latest browser versions, and updates were made a recommendation prior to further trialing. Low bandwidth or overloaded servers in schools led to problems with technical synchronisation, in which paired students became unpaired. In order to combat these technical obstacles a technical specification document was produced to outline the requirements for access to the assessment system. In response to teacher feedback and reoccurring issues, a troubleshooting document was also produced so that teachers had a technical guide on hand to address any issues during implementation.

Challenges

Australia's education system presents unique challenges to the development and testing of a system such as ATC21S. While the Australian system follows a national curriculum, the delivery, content, teaching and assessment of the curriculum content are in the remit of each state and territory government. This results in differences between the states and territories in what is taught in a given school year, and in how education is provided in terms of infrastructure and management.

Various issues were encountered while running the trials. The greater challenges lay in working with the different IT capacities in each school and the widely divergent understanding of IT systems among school principals and teachers. Because the ATC21S system had specific internet browsing requirements, the need for schools to download and install software – albeit free – presented an insurmountable barrier for some and no problem at all for others. Ensuring that schools had implemented the technical specifications before the trials commenced was another challenge. Sometimes the designated school contact for the NRC team was not very IT literate, so the technical specifications could be ignored as irrelevant until the day of the walk-through or the first trial, when the need for them to have been read and implemented became apparent. This happened often, regardless of the number and variety of reminders provided by the NRC team, highlighting the importance of training teachers and school leaders to give them a much better understanding of LDN-ICT. Infrastructure problems encountered were to do with slow or intermittent internet connectivity. This occurred most frequently in Western Australia, where teachers noted that the region-wide internet connection tended to be unstable during this period.

Some teacher pedagogical implications were recognised through this research and development. Firstly, an awareness of the importance of LDN-ICT and collaborative problem solving skills among teachers and school leaders needs to be achieved. Teachers who regularly adopt LDN-ICT and collaboration in their classrooms appear to have an advantage in the implementation of the assessment tasks. A teacher's classroom management skills can also impact quite strongly on the effectiveness of the assessments, particularly regarding the requirement for students to work using the chat box rather than speaking directly to one another. Some teachers found it easy to control their classes to this extent; others did not.

Secondly, an awareness of the 21st century constructs and skills needs to be achieved in order for them to be taught. Working with paper-based professional development modules, face-to-face workshops took place with two schools in Victoria. It was evident from the workshops that very little is required to set teachers on track. It is the quality of the information provided to them that is most important. Many of the skills evident in the assessment tasks are among those that teachers already observe in their classroom. However, teachers may be unfamiliar with the construct, terminology or presentation of the online assessments. By contextualising the skills, both online and offline, they become more aware of the characteristics of the skills and more ready to transfer their observations from one medium to another. Post-assessment debriefs between teachers and students appeared to be very useful, for both groups, in raising an explicit awareness of the skills being used and taught. After the assessment sessions, teachers would debrief with students by asking them open-ended questions, such as 'What skills did you and your partner use to work together efficiently?' and 'What was your process for solving the problem?' From this kind of exchange both teachers and students were able to identify which skills they are familiar with and use readily.

Lastly, if teachers are expected to implement 21st century skills in the classroom, further strategies will need to be identified and developed that are realistic to the time constraints of teachers. Working with teachers who have limited time presented challenges in gaining their commitment to the trials and in trying to ensure that the trials were run in a standardised manner. It seems a reasonable starting point for teachers to adapt the online tasks to offline classroom-based activities in which they can observe the same skills. An important concern for teachers was the need for additional time in their planning schedule and the development of their lesson plans to incorporate 21st century skills. A critical part of the workshop involved identifying how teachers could embed the new skills into the existing curriculum and therefore into already existing lessons, reducing the amount of preparation time required. Training for teachers, in some capacity, to familiarise them with the teaching of 21st century skills in their classroom, will be required if a major shift in this direction is to be seen. Further strategies for teaching 21st century skills need to be developed and the sharing of these between educators will be of critical importance. Considerable movement on this issue could be achieved at pre-service teacher level, supported by educational institutions.

Conclusion

The principal objective for Australia's National Research Coordinator's team was to test and validate assessment materials in Australian schools, and this was successfully achieved. The experience of the team in undertaking this project also provided useful feedback about the facilitators and the barriers that exist to the integration of LDN-ICT learning in Australian classrooms. Facilitating

(continued)

factors include the reforms that have been implemented by the Australian government via the DER, particularly the dedicated funding streams to develop the National Broadband Network and to provide ICT equipment to schools throughout Australia. These reforms have enabled Australian schools to embrace LDN-ICT as a valuable classroom tool and an important skillset for students to develop during their school years. The ATC21S project experience has also highlighted that programs designed to educate school principals and teachers in using and teaching LDN-ICT skills are necessary. Indeed, experiences in this project suggest that the wide disparity in LDN-ICT understanding that exists among school leaders and teachers is a significant barrier to the success of the Digital Education Revolution. The other major facilitator is the enthusiasm and dedication of teaching staff who are eager to embrace 21st century capabilities into their teaching, and willing to experiment with innovative methods to do so. These teachers and their school leadership have been overwhelmingly supportive of Australia's participation in ATC21S.

References

Australian Government. (2008). *Success through partnership. Achieving a national vision for ICT in schools. A strategy plan to guide the implementation of the Digital Education Revolution initiative and related initiatives*. Canberra: Commonwealth of Australia. Retrieved from http://www.deewr.gov.au

Australian Government. (2010a). *ICT Innovation Fund Guidelines, 2010–2012*. Canberra: Commonwealth of Australia. Retrieved from http://www.dewr.gov.au

Australian Government. (2010b). *Review of funding for schooling. Emerging issues paper.* Canberra: Commonwealth of Australia. Retrieved from http://www.deewr.gov.au

Australian Government. (2011). *School broadband connectivity survey 2010*. Canberra: Commonwealth of Australia. Retrieved from http://www.deewr.gov.au

Australian Government. (2012). Department of Education, Employment and Workplace Relations. Retrieved 08 Nov 2012, from http://www.deewr.gov.au

Australian Government. (n.d.). *Teaching teachers for the future – Building the ICTE capacity of pre-service teachers in Australian institutions*. Canberra: Commonwealth of Australia. http://www.deewr.gov.au

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

Delors, J. (1996). Learning: The treasure within. Report to UNESCO of the International Commission on Education for the twenty-first century. Paris: UNESCO Publishing. http://www.unesco.org/new/en/unesco/resources/online-materials/publications/unesdoc-database/

Education Services Australia. (n.d.). *ICT in everyday learning: Teacher online toolkit*. Canberra: Commonwealth of Australia. http://www.deewr.gov.au

Ministerial Council on Education Employment Training and Youth Affairs [MCEETYA]. (2008). *Melbourne declaration on educational goals for young Australians*. Melbourne: MCEETYA Retrieved from http://www.mceecdya.edu.au/verve/_resources/National_Declaration_on_the_Educational_Goals_for_Young_Australians.pdf

Chapter 9
Student and Teacher Responses to Collaborative Problem Solving and Learning Through Digital Networks in Singapore

Chew Leng Poon, Sean Tan, Horn Mun Cheah, Pik Yen Lim, and Hui Leng Ng

Abstract As a founder country of the ATC21S™ project, Singapore contributed actively in the task concept check, cognitive laboratories, pilot studies and field trials throughout the development of the ATC21S task prototypes. (The acronym ATC21S™ has been globally trademarked. For purposes of simplicity the acronym is presented throughout the chapter as ATC21S.) In all, 87 teachers/education officers and about 2,000 students aged 11, 13 and 15 from four elementary and eight secondary schools were involved in the project from 2010 to 2012. Besides capturing data on student performance in the tasks, Singapore researchers also interviewed teachers and students in order to better understand their attitudes toward the assessment of collaborative problem solving and learning through digital networks, and the challenges they faced in it. We found that teachers had to deal with "troubling" concepts in the new teaching and assessment paradigm – including the introduction of ambiguity into assessment tasks, tracking dynamic behaviours in collaborative settings, and the debate over content-rich and content-free assessment of 21st century competencies. Singapore students had fewer problems with learning through digital networks tools and skills than with skills of negotiation, group decision-making, communicating effectively to manage group dynamics and dealing with ambiguity and a less structured assessment environment. These lessons learned from the project provided useful pointers for Singapore as we enhance efforts in the teaching, learning and assessment of 21st century competencies in our schools.

C.L. Poon (✉) • S. Tan • H.M. Cheah • P.Y. Lim • H.L. Ng
Ministry of Education, Singapore, Singapore
e-mail: POON_Chew_Leng@moe.gov.sg

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_9

Context

Singapore is a small nation-state in South-east Asia, with about five million residents packed into all of 710 km^2 of land. It is home to a multi-ethnic and multi-lingual society of Chinese, Malays, Indians and people of many other ethnicities. Education is an important pillar of the Singapore social and economic architecture. The 2014 education budget of more than 11 billion Singapore dollars (about USD 9 billion) represents about 20 % of total government expenditure (Ministry of Finance, Singapore 2014) and is the second highest amount after expenditure on defence. This reflects the national priority given to education.

Formal schooling starts at age six in Primary 1 (equivalent to grade 1). Virtually all the half million students are enrolled in 357 publicly funded primary schools, secondary schools and pre-university institutions. Singapore has a national curriculum that provides equitable access to a broad and holistic education that includes the study of English, Mother Tongue Languages (such as Mandarin, Malay and Tamil), mathematics, the sciences, physical education, humanities and the arts. These subjects are complemented by co-curricular activities and community service programmes that develop life skills and socio-emotional competencies. All public schools base their teaching and learning programmes on the national curriculum and subject syllabuses. The subject syllabuses are reviewed regularly to ensure that they remain relevant for the future.

In 2009, Singapore developed a 21st century competency (21CC) framework to guide the development of its national curricula. The framework (Fig. 9.1) articulates the competencies that would enable students to grow into confident and concerned citizens with the necessary attributes and skills to learn continuously, work effectively in teams, exercise initiative, take risks and strive for excellence (Ministry of

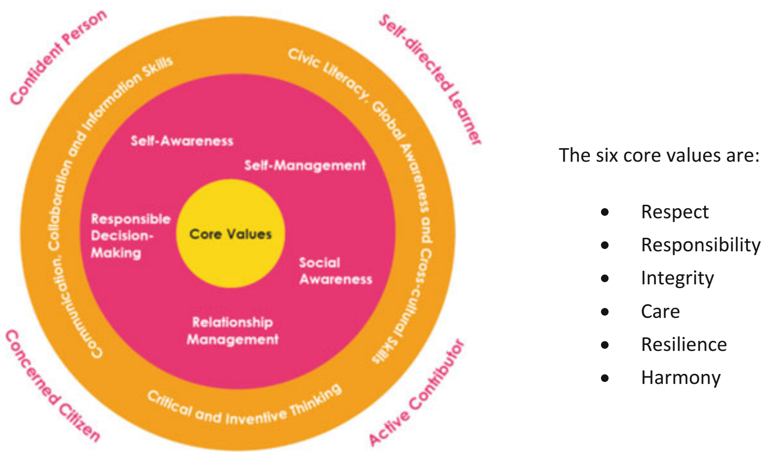


Fig. 9.1 Framework for 21st century competencies and student outcomes © Ministry of Education, Singapore (Reproduced with permission from the Singapore Ministry of Education)

Education, Singapore [MOE] 2010). The framework identifies three sets of enabling 21st century skills:

1. Civic literacy, global awareness and cross-cultural skills;
2. Critical and inventive thinking; and
3. Communication, collaboration and information skills.

The 21CC framework extends the work of *Thinking Schools, Learning Nation*, an education reform movement that began in 1997 to nurture a more thinking and inquiring mindset among Singapore students (Ng 2004; Sharpe and Gopinathan 2002). With the implementation of the 21CC framework, efforts were mounted to expand opportunities for all students to develop these competencies. For example, elementary schools enhanced their approaches to the teaching and learning of art and music to better develop creative capacities and personal, cultural and social identity in students (MOE 2011).

ATC21S Development in Singapore

It was during this period of heightened interest in the teaching, learning and assessing of 21CC that Singapore joined Australia, Finland, the United States, the University of Melbourne, and three international companies – Cisco, Intel and Microsoft – in founding the Assessment and Teaching of 21st Century Skills (ATC21S) project. Essentially, the ATC21S project sought answers to the following key questions (Griffin et al. 2012):

- What are the 21st century competencies?
- How can teachers teach them? How do students acquire them?
- How can students demonstrate them in measureable ways?

The ATC21S project in Singapore was spearheaded and funded by the Ministry of Education and drew in collaborators from the National Institute of Education, Nanyang Technological University. Twelve schools were also recruited as our partners in the project. Participating in the ATC21S project enabled Singapore to work with the international research community to establish assessment practices for 21CC, and more importantly, to find ways to automate such practices. These serve to complement and enhance our national efforts in the teaching, learning and assessment of 21CC. As the project developed intensively in the ensuing months, Singapore was heavily involved in the four critical phases of task concept check, cognitive laboratories, pilot studies and trials in the development of the task prototypes for the two sets of 21st century competencies – learning through digital networks – information and communication technology (LDN-ICT) and collaborative problem solving (CPS).

Prior to the actual data collection for the ATC21S project, it was hypothesised that the teachers and students involved in the study would have had limited exposure to the types of assessment tasks that were being developed in the project. Therefore, beyond the data that were specific to the ATC21S research questions, the Singapore team extended the scope to capture the teachers' and students' experience in their engagement with the tasks. The additional questions included: Did the teachers think that the tasks were assessing 21st century competencies? What aspects of the assessment and tasks did the teachers think would engage students, given the socio-cultural context of Singapore classrooms? What troubled or encouraged teachers when they thought about using the tasks for teaching, learning and assessment? Similarly, what aspects of the assessment and tasks did their students find engaging? And what aspects did students think were challenging?

In this chapter, we share the main findings and reflections from our inquiry into these questions within the context of the iterative process being used by the ATC21S tasks to further refine both the tasks and assessment practices. It is hoped that by sharing the voices of our teachers and students through the task development process, the international community can gain useful pointers in the collective efforts to assess 21CC attributes.

Method

In all, 87 teachers/education officers and a sample of about 2,000 students aged 11, 13 and 15 from four elementary and eight secondary schools in Singapore were involved in the four phases of task concept check, cognitive laboratories, pilot studies and trials (see Table 9.1). Griffin and Care (2015; Chap. 1) describes the processes involved in the four phases.

The 12 schools accepting the invitation to participate in this project had ensured that all teachers and students taking part had given their consent with the understanding that they could withdraw at any point if they wished. Table 9.2 summarises the sources of data for the current chapter.

Table 9.1 Number of ATC21S participants from Singapore

Task development phase	Number of schools	Number of student participants	Number of teacher/education officer participants
Task concept check	2 elementary and 2 secondary schools	–	32
Cognitive laboratory	2 elementary and 2 secondary schools	11 year-olds: 34	–
		13 year-olds: 25	
		15 year-olds: 13	
Pilot	2 elementary and 3 secondary schools	11 year-olds: 70	10
		13 year-olds: 66	
		15 year-olds: 98	
Trials	4 elementary and 8 secondary schools	11 year-olds: 232	64
		13 year-olds: 799	
		15 year-olds: 749	

Table 9.2 Sources of data for this chapter

Phase data was collected	Mode of data collection	Number of student/ teacher respondents
Task concept check	Focus group discussions and feedback form	Teachers: 32
Cognitive laboratory	Transcripts from students' think-aloud to understand their thought processes and how and why they responded in a particular manner in the tasks One-on-one post-task interview with students to elicit their views on the tasks	Students:
		11 year-olds: 25
		13 year-olds: 25
		15 year-olds: 9
Pilot	Post-task student survey and one-on-one interview with selected students to understand their attitudes toward and challenges faced in the learning and assessment of 21CC	Students:
		11 year-olds: 70
		13 year-olds: 66
		15 year-olds: 98
Trials	Post-task student survey and one-on-one interview with selected students and teachers to understand their attitudes toward and challenges faced in the learning and assessment of 21CC	Students:
		11 year-olds: 228
		13 year-olds: 738
		15 year-olds: 722
		Teachers: 3

Responses to the ATC21S Tasks

In our interviews of teachers and students, we asked for their views about using the ATC21S tasks for teaching, learning and assessment of 21CC. We also asked them to tell us which aspects of the tasks engaged, encouraged or challenged them. In this chapter, we have organised our teacher and student responses to the ATC21S tasks around six areas:

1. Relevance to the teaching, learning and assessment of 21CC
2. Engagement with tasks
3. Seeking meaning in content-rich and content-free tasks
4. Collaboration
5. Introducing ambiguity in tasks
6. Tools and technical issues

Relevance to the Teaching, Learning and Assessment of 21st Century Competencies

In general, the teachers saw the potential of using the ATC21S tasks for the teaching and learning of 21CC, barring customisation to suit local contexts. In reviewing the learning through digital networks tasks, teachers felt that the tasks were, in principle, aligned to the targeted constructs of functioning as consumers and producers of knowledge; and developing and sustaining social and intellectual capital. One elementary school teacher described how these tasks could help build the skills of "communication, reaching consensus when different ideas are there" which were "important, essential skills" that could support his school's 21CC program on "inventiveness". The CPS tasks were good exemplars of how to build in the need for students "to work together…and to really see strengths in the other person".

In previewing the tasks at the concept review stage, our teachers anticipated that the ICT-delivered tasks would likely appeal to their students whom they regard as "IT-savvy". Indeed our students, who were more exposed to pen-and-paper tests, were generally intrigued by the novelty of the tasks. However, 21st century competencies are not about the use of ICT alone. Instead, they are about the potential that judicial use of ICT can have on the transformation of key aspects of assessment, particularly the measurement of 21st century skills, such as metacognition, creativity and collaborative problem solving (Binkley et al. 2012), that are difficult to assess through pen-and-paper tests. The ATC21S tasks were good exemplars where there was clear leverage on ICT and social networking tools for online collaboration. However, one teacher observed that the ATC21S tasks have not made good use of the ICT platform to provide timely feedback to the students:

> I do not see any formative feedback given to students in the task. Students are not aware of whether they got the answer right or wrong, and how do they improve based on the mistakes they have made in solving the questions?

One student reiterated this observation:

> I received no feedback.

The teacher and student were not wrong in their expectations. Experts who worked on the ATC21S white papers have pointed to assessment innovation through advanced Web 2.0 technology that could tailor assessment and feedback to students even while they were working on the tasks (see Wilson et al. 2012). This is indeed an area of work-in-progress for the ATC21S team. In fact, data collected during the trials would provide a rich source of information to develop the learning analytics for the two sets of competencies, a step towards designing an automated system that could provide just-in-time probes to measure learning progress and to provide feedback. As one teacher put it, it would be excellent if students were learning "without realising that they are being taught". Such a system could also then address teachers' concern about the need to differentiate the tasks for students at different levels of proficiency in the 21CC.

Engagement with Tasks

Students were engaged by the dynamic and interactive nature of the tasks. For example, in one of the ATC21S tasks, students could "send" weights to each other to balance a beam. One student said: "I enjoyed the interactive tasks". Other students described the tasks as "fun, attractive, interesting". In fact, students' overall engagement level with the tasks was very high and many indicated that they would like to do similar tasks in school. From the post-task survey we conducted during the pilot and trials, close to 9 in 10 students agreed that the ATC21S tasks were interesting, more than 8 in 10 enjoyed solving the tasks together with their partner and more than 7 in 10 students preferred this mode of assessment to traditional pen-and-paper tests. These findings are consistent across all age groups.

Although students described many of the tasks as "fun", it did not mean that they found the tasks intellectually unchallenging. In fact, most of our students did not find the tasks easy, as one student summarised it:

> The task sets you thinking – puts thinking and analytical skills to use.

The appropriate level of intellectual challenge in a task plays a role in engaging students – studies have shown that students may become disengaged when tasks are not challenging enough (e.g., Hayes 2008). Conversely, task designers also need to guard against tasks that are perceived to be too challenging (Brophy 1987). Pitching the task appropriately for students is therefore an important consideration in engaging students in learning. It is not unreasonable to believe that this principle does apply to the learning of 21CC. In fact, the ICT platform that delivers the task can potentially be leveraged to differentiate task difficulty for different students.

Seeking Meaning in Content-Rich and Content-Free Tasks

How can 21st century competencies be best learned and assessed? Should they be embedded in content-rich tasks or in tasks that require very little disciplinary knowledge? This was one of the questions that the ATC21S team wanted to find some answers to. It therefore commissioned experts to develop both types of tasks. Different prototypes that drew on varying degrees of content knowledge were developed as contexts for measuring collaborative problem solving, ranging from content-rich tasks based on specific scientific or mathematical concepts (for example, Game of 20), to relatively "content-free" tasks that required students to recognise general patterns or rules from a real-life scenario (for example, Hot Chocolate).

Content-rich tasks generally sit well within most curricular frameworks that are designed along disciplinary lines, making it easier to identify the teachers who would teach the 21CC within disciplinary content. On the other hand, there are also concerns that the content within a task might alter proficiency estimates on the 21CC construct (Wilson et al. 2012).

Based solely on qualitative responses, we found that both teachers and students tend to struggle to find "purpose and meaning" when encountering content-free tasks. A student during a cognitive laboratory session thought aloud: "I don't get what you are supposed to do here. Asking to make a line appear but what is the line for? Purpose?" In another save/print task, a student said: "Don't understand what this is for… Since it is a tutorial to teach me something, just follow blindly what they are telling me." Teachers similarly searched for coherence and meaningfulness of the various activities, as illustrated in several teacher voices on the poetry task:

> What movie are you creating? What are the objectives? What is the feedback? What is the point of the exercise?
> It's testing the ICT skills, not the poem. They have the skills without learning anything.
> The students just do, don't know why.
> Meaning making is important.
> There is no link between the literary elements and the task. It's not really about this poem, what does it capture?

While it could be argued that these exemplified the cultural mind-set of teachers who were discipline-centric, or that these were unfamiliar experiences for teachers and students who tended to focus on assessment of content knowledge, it was also possible to understand from that context that pedagogically sound acquisition of content could actually help build the very competency measured in this task. As explained by a participating teacher:

> …there is room to build students' knowledge of literary elements and devices through the resources already made available in the scenario and exchange of views with partners to build new insights in the understanding of literary works. For example, while Singapore students may not understand the significance of the use of "Jim Crow" and the historical element of the Merry-go-round piece in America, actual exchange of views with other students can help them build this social understanding and capital.

Related to this is the question of whether the quality of thinking is integral to 21st century education. Beyond demonstrating the ability to use ICT and networking tools, is the quality of the output, in terms of what students write and create, and how they reason and justify their answers during the task, integral to 21st century teaching, learning and assessment? In terms of building intellectual capital, teachers expressed the view that the quality of the ideas and knowledge generated from the task was equally important to the 21st century skills that are being measured.

Perhaps the crux of the matter is not so much the question of whether the task is content-rich or content-free. Regardless of whether it is anchored in a discipline or not, a more important issue is whether students could find meaning and authenticity when working on the tasks. The tasks could be scaffolded to build up towards meaningful goals, as meaning making is an important aspect of learning (Perkins 2009). As a 15-year old commented: "I was bothered from the beginning on what was the aim of the activity". There is therefore room to consider this perspective of meaning making for teachers and students in the design of the tasks.

Collaboration

There was considerable deliberation within the ATC21S community on linking the design of the tasks to the accurate assessment of collaboration. The key debate revolved around the balance in the provision of symmetric and asymmetric access to information that would facilitate meaningful collaboration. Each approach seemed likely to solicit different aspects of collaboration skills and provide avenues for the development of these different dimensions.

The developers for the learning through digital networks tasks believed that the same information should be available to all players in a collaborative setting, while the developers for the collaborative problem solving tasks ensured collaboration by giving different collaborators access to different sets of information. We believe that the ideal scenario would be a good balance between the two across a comprehensive range of tasks. We made some observations on how our students collaborated based on the current tasks. Our students responded to "collaboration" tasks in several ways – (i) one student worked out the solution while the partner passively agreed to the solution; (ii) one student worked out the solution while the partner verified the answer; (iii) the two students discussed the problem and worked out their solutions collaboratively. Our students tended not to collaborate on tasks where they and their partners were presented with the same information. Collaboration was not seen to be critical in accomplishing the task, as students could attempt the questions without working with others. In other words, students saw the problem solution as the larger goal and sidelined the collaboration when they perceived that they could solve the problem independently.

During the interviews following the cognitive laboratory, students explained that the different pace at which their partners completed each task affected the quality of collaboration (each group had between three to four collaborators during the cognitive laboratory for learning through networks). One student reached a collaborative task in Arctic Trek much earlier than her team-mates. She initiated a chat but later decided to skip that collaborative chat and move on to the next task after having no response from her partners, none of whom had arrived at this part of the task yet. She said: "It's frustrating when my friends don't answer back".

One aspect that students liked about the collaboration tasks was the opportunity for negotiation and decision-making among members who collaborate. Our students liked the opportunity of listening to "conflicting opinions", of "giving my own opinions" and of "seeing friends' thoughts in Webspiration (the poetry task)". The students did not often encounter such opportunities in their regular classrooms, as one student commented: "This kind of online negotiation, discussion and decision-making is not common during school work." Indeed, learning how to negotiate and make decisions in a team are important competencies to develop.

One of the things that we are curious to find out from the ATC21S data is the impact of student ability on collaboration outcomes. For example, we observed that when a very academically able student was paired with an academically weaker student, there appeared to be domineering behaviour by the more able and articulate student to take the lead in the task. We are not sure if this observation was an isolated event or that prior ability of students could have an impact on the levels and quality of collaboration. Further analysis of the actual ATC21S data collected will be helpful in answering this important question.

In one of the interviews, one student told us: "I find it very difficult to answer and work with others through on-line collaboration". It made us wonder if measures of collaboration would also depend on the modality (e.g., using online chat versus audio chat) and platforms (e.g., offline versus online) of collaboration. This would be an interesting follow-up study that could help improve the design of 21CC collaborative tasks.

Introducing Ambiguity in Tasks

One consistent observation made by our students and teachers was that the ATC21S tasks were less well-defined than the tasks in traditional tests. Students generally had to figure out the problem they had to solve and the approaches they needed to take to solve it. For example, in the Warehouse task, where the objective is to secure a warehouse by correctly positioning security cameras, collaborators are presented with different information – one can see and place cameras while the other sees only yellow beams (which show the coverage of the cameras placed) but not the cameras. Students need to realise that they have control of different parts of the problem, and in this case, figure out what the yellow beams represent. In the cognitive laboratory, we often heard students say:

> What is the question?
> What am I supposed to do here?
> How am I supposed to do this?
> The instructions are not clear.

One student said there was "not enough information given in the tasks on how to use the relevant websites to answer the questions". The younger elementary students described the task instructions as "confusing". The students were searching for "clear instructions," as one student put it – something they were more used to in the tests they have usually encountered. Students said they have to "figure out on my own", "spend a lot of time to figure out what was expected" and "when I found it, which course of action I should take, then much clearer. The activity became much more understandable."

One teacher made a keen observation about the tasks: "Introduction of ambiguity into the tasks creates space for students to think, inquire and collaborate". Indeed, many of the ATC21S tasks were deliberately designed to be less structured and more ambiguous than traditional assessments, and to provide collaborators with different 'views' of the problem. This created space for students to think about what the problem was and to collaborate and work out different ways to go about solving it. This is an attempt to better reflect the reality of 21st century contexts where problems are ill-defined, information/expertise resides with different sources and solutions are neither immediately obvious nor straightforward (NRC 2011).

While we support the introduction of ambiguity into the 21CC tasks, we suggest that task designers pay more attention to how the goals and instructions of the task are crafted, providing clarity where ambiguity is not intended so as to reduce confusion and not to discourage students from continuing or even getting started with the tasks. For instance, a 15-year old student commented during the cognitive laboratory: "Reflect on the poem. What are they asking in this question? Is it evaluation of poem or usefulness of poem?" The extent of ambiguity introduced should perhaps also take into consideration the age and proficiency level of the students. Younger students or students starting from a lower level of proficiency might require more scaffolding and feedback to help them stay engaged with the tasks instead of giving up because they were too "confused" or even discouraged (see for example Kirschner et al. 2006). Already, some of the ATC21S tasks are good exemplars of problems for which scaffolding has been built into the design. For example, in a task based on a numbers game, students first play against the computer on their own to understand how the game is played, before they move on to work with their partner to play against the computer. In another task, students can review the outcome of the previous page to understand how the puzzle is solved and apply their learning to the problem on the next page (which is built on the previous puzzle). In other words, the problems are designed to be increasingly less well-defined as the students progress through the task. These are important task design considerations, especially when working with different learners.

Tools and Technical Issues

Choice of online tools and support for technical issues are important in task design to minimise confounding factors in measuring student proficiency levels in the 21CC. The design of tools that students need to use in tasks should be user-friendly and intuitive so that students are not hampered by the complexity of the tools in progressing within the task. In the poetry task, students across all age groups had difficulty using the mind mapping tools – many found it confusing and most challenging. One student who skipped the page said during the post-task interview that he knew what a mind map was and how to draw one, but he did not know how to use the tools in the task to do that. Another student thought aloud: "Cannot seem to do anything… This page is confusing, how to use this? Not sure what to do." The student later explained that the task did not give specific instructions on how to use the tools and the tools were not intuitive. For example, the pencil icon did not visually tell a user that it could be used to create a label and a link.

Consequently, the interface design and interactivity of the tools could be aligned with common software to avoid confusion. For example, students were generally proficient and comfortable with using the chat function in the tasks, which they found to be similar to chatting on Facebook. However, students felt that they should not be restricted to using the chat function to communicate, and preferred to use social networking tools that were more common and authentic, e.g., Facebook, MSN messenger, even mobile phone.

There were some technical issues that surfaced during the trials, such as server failure, long loading times between pages of the tasks, system hanging, and technical bugs in the assessment software, that could have an impact on the delivery of the assessment as well as on students' motivation and engagement with the activities. While some of these problems could not have been prevented, the lesson learnt is that it is crucial to pre-empt infrastructure problems and conduct comprehensive compatibility testing with the operating platform prior to the implementation of any computer-based assessment.

Challenges

Assessment practices tend to drive learning behaviours and trigger pedagogical responses. For an essentially centralised education system that has achieved strong recognition on the international stage for its students' academic attainment, a fundamental change in what is being assessed requires substantial efforts. The changes needed are not confined to the professional space but will likely involve social and economic shifts. This represents a key challenge in introducing the assessment of 21st Century skills at the system level.

At the practice level, the tools developed under ATC21S are at best at the nascent stage of an exciting journey. While they point strongly to a reasonable way forward in attempting to automate such assessments, there is still substantial development ahead. For our context, the interpretation of the actions of the students when attempting the tasks will likely need refinement, as such interpretations need to take into account cultural differences amongst others. On top of this, the discourse of the students which is captured could usefully be analysed. This layer of interactions will add richness to the assessment of the students' 21st Century skills, without which the picture will not be complete. As our understanding of 21st Century skills and their assessment increases, curricula and pedagogical adjustments will need to keep pace to make the collective efforts meaningful and effective.

On the technical level, the key challenge is the ability of the computing system to capture and accurately interpret discourse (including both verbal and non-verbal). Given that natural language processing is still fairly far away from being ideal, and that the ability to interpret human factors such as facial expressions and body language is only at the development stage, a fully automated system for assessing 21st Century skills will not be easy to achieve. A blended approach, at least till the technologies catch up, with teachers being effectively supported by systems that can take up the bulk of the assessment load, could be a reasonable interim practice that helps support ATC21S and similar work.

Conclusion

Most countries, including Singapore, have minimally developed a curricular framework or policy on building students' 21st century competencies. While there is some extent of agreement on what constitutes 21st century competencies (see for example Voogt and Roblin 2012), there has been less clarity and

(continued)

agreement on how to go about teaching and measuring these competencies. The methods and approaches to doing so cannot be easily extended from current knowledge and the current capabilities of educational psychometrics. They require re-framing as well as new skills and knowledge in the development of tasks and the teaching and measuring competencies. In this sense, the ATC21S international project has made some headway in prototyping what is possible. We are deeply grateful for the honest and constructive feedback that our Singapore teachers and students have given us throughout this project. We take their feedback seriously, as they are the key actors in the teaching and learning of 21st Century skills. We are cognizant of the fact that there is more to be done and look forward to contributing to this important area of work with our partners.

References

Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Eicci, 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.

Brophy, J. (1987). Synthesis of research on strategies for motivating students to learn. *Educational Leadership, 45*(2), 40–48.

Griffin, P., Care, E., & McGaw, B. (2012). The changing role of education and schools. In P. Griffin, B. McGaw, & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 1–15). Dordrecht: Springer.

Hayes, D. (2008). *Foundations of primary teaching* (4th ed.). London: Routledge.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*, 75–86.

Ministry of Education, Singapore. (2010). *Nurturing our young for the future: Competencies for the 21st century*. Singapore: Ministry of Education. Retrieved 21 Jan 2013, from http://www.moe.gov.sg/committee-of-supply-debate/files/nurturing-our-young.pdf

Ministry of Education, Singapore. (2011). *Education in Singapore: Findings from international benchmarking studies*. Singapore: Ministry of Education.

Ministry of Finance, Singapore. (2014). *Singapore Budget 2014 – revenue and expenditure estimates: Government expenditure*. Singapore: Ministry of Finance. Retrieved 28 May 2014, from http://www.singaporebudget.gov.sg/data/budget_2014/download/5%20Government%20Expenditure.pdf

National Research Council. (2011). Assessing 21st century skills: Summary of a workshop. In J. A. Koenig & Rapporteur. Committee on the Assessment of 21st Century Skills (Eds.), *Board on testing and assessment, division of behavioral and social sciences and education*. Washington, DC: The National Academies Press.

Ng, P. T. (2004). Students' perception of change in the Singapore education system. *Educational Research for Policy and Practice, 3*, 77–92.

Perkins, D. (2009). *Making learning whole*. San Francisco: Jossey-Bass.

Sharpe, L., & Gopinathan, S. (2002). After effectiveness: New directions in the Singapore school system? *Journal of Education Policy, 17*, 151–166.

Voogt, J., & Roblin, N. P. (2012). A comparative analysis of international frameworks for 21st century competencies: Implications for national curriculum policies. *Journal of Curriculum Studies, 44*, 299–321.

Wilson, M., Bejar, I., Scalise, K., Templin, J., Wiliam, D., & Irribarra, T. (2012). Perspectives on methodological issues. In P. Griffin, B. McGaw, & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 67–141). Dordrecht: Springer.

Chapter 10
Introducing Assessment Tools for 21st Century Skills in Finland

Arto K. Ahonen and Marja Kankaanranta

Abstract The Finnish national interest in the enhancement of 21st century skills has highlighted a need for and interest in developing tools and methods for teaching and assessing such skills. In this chapter, we present and analyze the development process of online assessment tools in the Assessment and Teaching of 21st Century Skills study (ATC21S™) from the Finnish perspective. (The acronym ATC21S™ has been globally trademarked. For purposes of simplicity the acronym is presented throughout the chapter as ATC21S.) The development process was implemented according to the guidelines of the international project through four phases, namely concept checks, cognitive laboratories, pilot studies and trials. These phases are analyzed from the student and teacher perspective. This chapter presents the experiences, possibilities and challenges of introducing and developing the assessment tasks for 21st century skills in the Finnish comprehensive schools across the different phases of the study. An essential element of the process was the need to translate and localize the tasks to the Finnish contexts and language. The chapter also discusses ideas for the further development of such tasks towards more collaborative research design.

Context

During recent years, the role of 21st century skills in teaching and learning has received a lot of attention at various levels of the Finnish educational system. A national expert panel indicated that Finnish school leaders and politicians agree that the country needs to make 21st century skills more prominent in its schools (Salo et al. 2011). At the policy level, the objectives in the new Basic Education Act for the year 2020 carefully define 21st century skills in accordance with the framework created in the international Assessment and Teaching of 21st Century Skills study (Ministry of Education 2012; Binkley et al. 2012). There seems to be a common

A.K. Ahonen (✉) • M. Kankaanranta
University of Jyväskylä, Jyväskylä, Finland
e-mail: arto.k.ahonen@jyu.fi

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_10

understanding that the development of our school system necessitates the better embedding of 21st century learning as well as the design of new assessment tools and methods in order to provide citizens with better capabilities to participate in the “knowledge society” (Krokfors et al. 2010; Norrena et al. 2012).

At the same time, there have also been debates suggesting that Finnish schools are at a crossroads and must decide whether they will keep up with developments in other sectors of life and society or follow a separate path (e.g. Pohjola 2011; Vähähyyppä and Mikama 2010; Välijärvi 2011). Moreover, research evaluating Finnish school curriculum implementation has shown that even though 21st century skills are well recognized and referred to in the curriculum, they do not yet have a role in everyday school practices and are mostly left behind in regular teaching and learning (Holappa 2007; Kankaanranta and Puhakka 2008; Kartovaara 2009; Siekkinen and Saastamoinen 2010). The majority of the 21st century skills and competencies are embedded in cross-curricular themes in the Finnish National School Curriculum. Although these themes are regarded as central at curricular level, putting them into practice is not always easy. Kartovaara (2009) indicated that school leaders in Finnish comprehensive schools do not consider these themes to be well established in their teaching; also, they find them difficult to teach. Moreover, there are differences between schools in the degree to which they adopt 21st century skills in their teaching and learning programmes. A recent study conducted by the Finnish National Board of Education found that students’ knowledge of cross-curricular themes was good but that their attitudes toward them needed improvement (Lipponen 2012). Teachers think that cross-curricular themes do not have a clear enough role in the school curriculum and they are therefore often absent from their teaching (Niemi 2012).

ATC21S Development in Finland

The national interest in the enhancement of 21st century learning has also brought about a need for and interest in developing tools and methods for teaching and assessing these skills. This led Finland to join in 2009, as one of the founder countries, the Assessment and Teaching of 21st Century Skills project. Finnish participation was justified by Finland’s Minister of Education and Science as follows (Ministry of Education 2009):

> From the Finnish perspective, the project offers an opportunity to develop teaching methods and learning environments and to enhance pupils’ creativity, social skills, innovativeness and problem solving skills.

The project partner in Finland was the Ministry of Education and Culture (member of the executive board Dr. Sakari Karjalainen) and the project was conducted by the Finnish Institute for Educational Research in the University of Jyväskylä (National Project Managers professor Marja Kankaanranta and researcher Dr. Arto K. Ahonen).

Method

The main international objective set for participating countries was to test and validate ATC21S assessment materials in schools through four phases, namely concept check, cognitive laboratories, pilot study and trials. The phases were conducted in order to gain information about the suitability and quality of the tasks for students of school age. The schedule for conducting the different research phases was dependent on the progress of the concept and task development. Finland faced the additional challenge of translating and localizing the assessment scenarios and other relevant materials into the Finnish language. Table 10.1 presents the ATC21S research design in Finland. The evaluation of the two task sets – Collaborative Problem Solving (CPS) and Learning through Digital Networks – Information and Communication Technology (LDN-ICT) – proceeded through separate timelines, determined by the availability of the concepts and scenarios.

The concept checking of LDN-ICT literacy tasks took place in two schools with 13 teachers at the beginning of December 2010. At the first school, eight teachers participated in a 3-h session. In the other school there were three separate 2-h sessions for seven teachers. The participating teachers were classroom teachers for grades 5–6 and subject teachers for grades 7–9. All the teachers evaluated three available scenarios and they responded to three questions after viewing the tasks. The concept checking of collaborative problem solving (CPS) tasks was carried out with four teachers in one school in May 2011. This was implemented with the English version of the tasks since Finnish versions were not yet available.

For the cognitive laboratory, the tasks were translated into Finnish. Cognitive laboratories of CPS tasks took place through two phases in two schools with 28 students: in May 2011 for the first set of tasks and in October 2011 for the second set of tasks. The cognitive laboratories for LDN-ICT tasks were conducted in one school with 12 students in October 2011. A group of four students from each of the three chosen age groups – 11, 13 and 15 year-olds – participated. The pilot and trial studies were conducted only for the CPS tasks due to server problems with the LDN-ICT tasks. Participation details are shown in Table 10.1.

The trial study was carried out in six schools with a total of 520 students. Interviews with students were conducted on a volunteer basis with 14 students in groups of 2–4. Students were asked for their opinions of the tasks (e.g. general opinion, difficulty level, enjoyment level), their familiarity with such tasks, and suggestions for further development of the tasks. In addition we carried out two expert evaluations of the CPS tasks.

Localization and Translation

The process of translating and localizing the ATC21S assessment tasks into the Finnish language and context continued throughout the research schedule. This resulted in several new versions of the scenarios and raised various problems in the

Table 10.1 ATC21s research design in Finland

	CPS tasks			LDN-ICT tasks		
Project phase	Method/data	Participants	Schedule	Method/data	Participants	Schedule
Concept checking	Short teacher questionnaire	1 school: 4 teachers	May 2011	Short teacher questionnaire	2 schools: 13 teachers	December 2010
Cognitive laboratories	Guided observation, think-aloud sessions, research diary	2 schools: 24 students	May & October 2011	Research diary	1 school: 12 students	October 2011
Pilot study	Participatory observation	1 school/72 students: 5th grade 24 students, 7th 16, 9th 32			*Not conducted*	
Trial study	Student interview, research diary	6 schools/520 students	November 2011, May 2012		*Not conducted*	

different phases of the schedule. The concept checking was carried out with the English versions of the tasks. Thus, the actual task translation started with the cognitive laboratories. The translation process was strongly affected by the schedules of the three task developers as well as by the diverse solutions they provided for embedding Finnish texts on the tasks. This required intense collaboration with the developers. The main steps in translating the assessment tasks were the following: basic translation, either by a specialist translation company or the researcher; moving the translated texts to the electronic version; and further localization of the tasks. There were differences in regard to the demands of the translation procedures across the tasks in order to make them more suitable for the Finnish school context.

The translation and localization process of the CPS tasks was relatively straightforward. The translations were sent as Word documents to the University of Melbourne to be placed in the server and a Finnish player portal was developed. For the ICT tasks, the developer, UC Berkeley's BEAR Center, provided an online translation platform, which enabled the translations to be made directly to the task scenarios. The translation of the ICT tasks was complicated, especially in regard to the use of external websites related to the tasks. For the purposes of the research phases, the decision was made that the external websites and services linked into the tasks would be left in English. The students were provided with help podcasts with Finnish subtitles. The questions of whether and how external content should be translated remain to be answered.

The cognitive laboratories showed the need for further localization, due to linked pages with language-intensive content in the tasks. Even though Finnish students are familiar with using English as an operational language on the Internet, it cannot be assumed that the tasks can successfully be completed by them in English. When the assessment portal is connected to an external website, it is challenging to provide a substitute that matches the content in another language and culture. The machine translators do not yet work well enough to provide content that matches the English versions. However, specific solutions for the use of external content can be tailored for the purpose, such as whether they are developed for large-scale comparative purposes or more for student-centered use as learning tasks.

In Finland, the tasks were initially translated by researchers and checked by professional translators. Back-translation was not conducted. Notwithstanding this shortcoming, the students appeared to respond to and engage with the tasks in Finnish in a manner similar to that observed among the English-speaking countries, and to that extent we have some indications of the face validity of the translated materials.

Teacher and Student Responses to the ATC21S Tasks

Concept Checking

The implementation of the concept checking sessions in Finland raised some problems related to the overall concept checking procedure, the tasks to be checked, schools' ICT access and the use of English versions of the tasks. The teachers were able to operate in English, as was presumed. However, this naturally added to the demands of the concept checking and it remains unclear what effect this has on the results of the evaluation, e.g. how accurately teachers can evaluate the questions or activities in relation to the student capabilities when the tasks are in another language. The concept checking of technology-based tasks necessitates also that certain access requirements are met at the schools. The online collaboration requires rather fast and robust connections. In spite of these difficulties, the teachers had a positive attitude towards the concepts and some teachers even continued with the work after the sessions.

The concept check sessions were preceded by a researchers' visit to the school, where the ATC21S project and 21st century skills framework were briefly presented to the participating teachers. For the LDN-ICT tasks, teachers were asked to respond to three questions concerning task scenarios. Firstly, the concept checking gave teacher-based information on the skills and competencies that the scenarios target. This can indicate how the scenarios correspond to the goals based on which the scenarios were developed. For the LDN-ICT tasks, the question is whether the scenarios relate to LDN-ICT literacy skills. Teachers named only rather restricted areas of LDN-ICT literacy, such as basic ICT skills, the understanding and ability to use multiple tools, searching information on the web, ability to sign into web based services, and uploading and restoring digital information. The more complex LDN-ICT literacy skills were not mentioned in teacher responses. The Finnish data indicated that the concept checking could provide interesting insights into how teachers conceptualize and understand 21st century skills in general. The data also provided scenario-specific information about the spectrum of skills that teachers think can be learned and assessed through the scenarios.

Secondly, concept checking provided teacher ratings on the applicability of the tasks, scenarios, and activities at different school levels. It provided information regarding the difficulty of the tasks and related age suggestions. In general, Finnish teachers evaluated the scenarios to be too difficult for 11-year-old students. There was, of course, variation between activities and questions. Generally, activities such as uploading and restoring information, following instructions and some activities like the graffiti wall and movie task were assessed as difficult for 11-year-old students. Other tasks, such as chatting and using a rating system were assessed as easy. Chatting was regarded as something the students could do in English.

> The movie task was difficult for 11 year-olds, but also probably a task they'd like. There should be also some directions what to film, when making the movie, not only to film something. It could be like just reading of the poem and filming it. (Teacher, 5th grade)

The concepts were checked in English, which caused problems for some teachers: naturally, it is more difficult to evaluate the difficulty level if the concepts are not in the language to be utilized in the final tasks.

Thirdly, the concept checking offered teacher-based insights into the use of tasks in a school setting and also into student perspectives, e.g. what kinds of tasks and activities teachers believe that students would enjoy. For example, one teacher mentioned that online chatting is familiar for students, though a novel means of communication for many teachers. Teachers were inspired to raise issues related to scenario ideas – their further development, their relation to student work and motivation, and also more practical issues. For the further development of the teacher questions, we suggest the addition of a specific question, which asks teachers to think over the applicability of the scenarios in school settings.

Cognitive Laboratory

The cognitive laboratory provided us the first experience of these kinds of tasks in use at schools, with real students in real contexts. Through this phase we could monitor the suitability and functionality of the tasks as well as get ourselves prepared for larger-scale pilots and trials. In this way the cognitive laboratories provided teachers, students and the research team with the opportunity to practice the tasks.

In the cognitive laboratories of CPS tasks, groups of four students were introduced to the assessment tasks and asked to complete them. Teachers from participating schools were intensively involved in this phase and, along with the researcher, they helped collect the observation data. During all the sessions we had one teacher or school assistant observing each student, so we received very precise information on the process and the actions of the students during the completion of the tasks.

The students were prompted to "think aloud" when they were completing the tasks. They had microphones and voice recording programs running, but it soon became clear that the students did not talk aloud at all. The recording equipment provided no useful data, perhaps because thinking aloud or speaking to yourself while someone is sitting beside you is not a very natural thing to do. Perhaps, also, the tasks were absorbing or demanding enough to take up all the students' attention. At the beginning, the students were anxious and even frustrated, because the experience was so unfamiliar to them. Working in a pair online, formulating collaboration in the chat box, and working with difficult tasks with strategically unclear directions, were just some of the challenges they faced. Also the students were advised to try the tasks without requesting help from anyone other than their partner.

The main part of the data from the sessions consists of the observation sheets (altogether 260 pages) completed by teachers when observing their students completing the tasks. The work of the students was also captured via screen capturing programs, and their actions on the computer could be checked and analyzed afterwards, along with the discussions. The observation sheet consisted of 20 elements across participation, task regulation, social regulation, perspective taking, and knowledge building. There was a sheet for each page of the tasks, and the observers filled them in simultaneously. From the observation data it can be seen that, when completing the tasks, the students commonly showed engagement and interacted with their partners through the messaging tool of the task. The higher-order elements such as reasoning were seldom observed.

The cognitive laboratory phase included an optional student interview. Our aim was to complete our understanding about the tasks, and their applicability, from the student perspective. The student interviews provided interesting input from the students for the further development of the tasks. Most of the students found the tasks challenging or difficult. This was in line with the results from teachers' task evaluation in the concept checking phase. The main difficulties lay in problems with understanding the directions and in collaboration with the partner. Many of the students' direct suggestions for changes to the tasks focused on the development of more simple tasks and clearer directions.

> The tasks were really challenging, you needed to think about them for a while, but there were also a couple of tasks that were rather easy. (Girl, 7th grade)
>
> I think the tasks were not so difficult, but it was just not possible to figure out what I was supposed to do. The directions were so unclear. (Girl, 5th grade)
>
> [...]And if you had a different view from your partner, and the partner had some extra material, then if the partner was not active, you just could not solve the tasks by any means. (Girl, 5th grade)

These comments confirm that the tasks are setting a level of challenge that requires the employment or development of new skills. The unfamiliarity of the tasks was further underlined by requests for more action and game-like movements. There were also requests for improvements in the chat box. The interviewed students also reported that they did not have prior experience of such school tasks.

> They could have involved more action. Something else than just moving the mouse. (Girl, 5th grade)
>
> Game-like, and some background music also. (Girl, 5th grade)

Pilot and Trial Study

In Finland, the pilot and trial study were conducted only for the CPS tasks. In the pilot study, the main focus was on solving the practical issues during the task completion procedure. This was the first opportunity to test the assessment tool with a whole class. The main focus of the pilot was on the administration and login procedures during the task completion. The pilot study revealed several minor problems. Most of them concerned technical issues such as logins, network connections, task completion, and the pairing of students. The technical issues required a lot of attention but did not detract from task completion with strong commitment from teachers and students.

It was intended that the teachers administer the field trials in the schools, but it became apparent that the presence of the researcher was needed in most of the schools. The whole procedure of using online tasks was new for all the schools and the task administration was quite complex. We therefore administered the trial study session with the researcher attending onsite at schools. Only in one school, where the principal was involved in the process from the beginning and conducted the first few sessions along with the researcher, was it possible to complete the trial studies without the researcher. In most of the sessions there was also a teacher present during the task completion, which was very useful. The students asked questions and needed help regularly, and their own teacher was the most convenient person for them to consult.

Based on observations from the trial sessions, it was possible to identify a common pattern in the behavior of the participating student groups. The students were confused at the beginning of the sessions, tended to discuss with each other and ask questions, but towards the end reached some level of enthusiasm and worked seriously on completing the tasks. The following excerpts from research diaries accurately describe the sessions.

> The 9th graders were all at the same time in the trial sessions, so we worked at two separate computer labs. The atmosphere was at the beginning confused, but changed rather quickly into positive atmosphere of working and you could recognize even some elements of flow in the class. Some tasks seemed to really challenge the students to think and use their capacity. All students were really working enthusiastically and trying their best. (9th grade trial, 3rd Nov 2011)
>
> During the practice task there was a lot of confusion and questions in the air. But when they reached the first tasks the working seemed to get better. Still during the task completion there was some moving, confusion and talking going on. Pairing of the students seems to cause most of the confusion; which is paired with who and why. Towards the end of the session it cooled down and the working got going. (9th grade trial, 11th Nov 2011)
>
> The atmosphere was rather restless, because all the time some pair was dropped off or needed some kind of technical help. Only for a little moment at the end of the 90 minutes session there were calm and nice atmosphere with active working. It can be seen that 50 students from two different classes are too many to handle in one session, even with two teachers helping. The feeling was left a bit unclear and at the first time I felt that I was not sure whether the tasks are working and whether the students understood what they were about. (5th grade trial, 21st Nov 2011)

During the practice task the students were allowed to discuss and ask questions. Usually the session was broken off after the practice task and the students were told to try to manage without discussion and questions and to limit their communication to the chat function in the tasks. After a while it was possible to recognize a feeling of enthusiasm when the students were completing the tasks. This period of silent and concentrated work varied but usually lasted at least half an hour. The students seemed to improve their working throughout the session, and in their peer and self-assessments they usually demonstrated a good awareness of their level of working.

Challenges

Translating learning or assessment tasks from another language is a demanding process and requires several phases to verify that the task corresponds to the original version. The tasks also need to be carefully localized to the local language and culture. For example, the translation process in the PISA 2009 study required separate processes in translating and verifying the tasks in the local context (Arffman 2012). According to Arffman (2012) there are several procedures and practices that need to be taken into account when translating international achievement studies. The tasks should be developed to be as translatable as possible, the translators should have sufficient knowledge of the task contents, the translation process should have clear goals and guidelines, there should be enough time for revision and verification, and the verifiers need to have enough knowledge about source languages, subject matters and familiarity with testing (Arffman 2012). Even when translation procedures are carefully followed and all these steps carried through, it is likely that full equivalence is never attained (see Chesterman 1997; Grisay et al. 2009).

The concept checking of the task scenarios was quite time-consuming and required some intense work from teachers, who struggled to assess all the scenarios carefully in one session. Some teachers commented that they found the work tiresome. The level of intensity required by the tasks could usefully be taken into account in further analysis of the workload produced by the tasks for students. The participating teachers reported that they did not have previous experience related to the concept they were asked to check and, for most of them, the procedure of concept checking was itself a novel experience. Yet they warmed to the assignment and quickly started to evaluate the suitability and usability of the assessment tasks. The online collaboration regarding the LDN-ICT tasks also presented some challenges for the teachers. Some of them managed to chat online easily and found it useful, while others seemed to face problems. It also seemed that more than one session was needed for the teachers to work out what the task was about. Many of the teachers logged on themselves later and commented more then.

In this form, the cognitive laboratory as a method was intensive and time-consuming with heavy observation tasks and programs. However, it has potential for better understanding of the data needs and their usability for national purposes, as well as for greater feedback to schools and national researchers. This phase could also include other methods for gaining student and teacher evaluation of the task completion and use at schools.

In the trial study, it appeared that the school leaders were the key players in the recruiting of schools. The bigger the school, the more difficult was the practical organization of the trial sessions. In small schools, where only one or two groups of each grade participated, fitting the booking of computer labs and weekly schedule worked out rather easily. In bigger schools with six groups in each grade level, the arrangement of free time and space was difficult to organize, and it detracted from the completion of the study. One school pulled out of the study at this stage, when it learned how much time would be required to complete the study for several groups.

From the concept checking we realized that teacher participation acted as an important and thorough means of involving teachers in the whole research process. Through it, they internalized the meaning of the skills and tasks and were well prepared for the phases that followed. Finnish teachers were interested in – some even enthusiastic about – the novel concepts related to the assessment of 21st century skills. They recognized the need for such tasks and saw their potential as learning and assessment tools. It is worth asking whether teachers who participate only in the later phases of the research can become as involved and enthusiastic as those who participate in the concept checking. Our experiences suggest that the project is at its best when the same teachers and schools are involved through the whole development process.

The implementation of the concept checking sessions at schools demonstrated that the procedure should have a clear schedule, include added value for the participants – for supporting involvement of schools and teachers – and careful mapping of access issues at schools to ensure that access is suitable for concept checking. Our experiences indicate that there is a need for clear guidelines on how the teachers are introduced to the whole task development process and to the 21st century skills. The concept checking phase works well as the first step, motivating and inspiring participation in the other phases. It also acts as a deep introduction to the activities that follow, as well as to the principles of learning and assessment of 21st century skills. The more information teachers are given, the more involved they feel. This brings us to the next important issue, namely what is the added value of participation for the teachers.

Cognitive laboratory was an intensive, challenging, and rewarding phase of the study. The method produces a lot of valuable information about the students' abilities to solve the problems and about their behavior while engaged in the tasks. The fact that the teachers acted in the researcher's role also has implications for further research. The teachers are a valuable source of information about the administration of the study. A large database of teachers' and researchers' observations made during the task completion is now available. The Finnish data covers very interesting research themes concerning possible cultural differences or similarities in the task completion or communication strategies, as well as information relevant to many other interesting questions for further research.

Logistically, the pilot study should be the trial study on a smaller scale, i.e. with a smaller number of participants but with similar data collection methods. Conducting pilots in schools already known to the researcher was a clear advantage. It helped to avoid unexpected surprises, and provided a familiar computer lab and collaboration with teachers who had been involved with the tasks since the concept checking phase. With this experience, we were better able to introduce the tasks for whole classes of new students with confidence in the existing infrastructure, giving both teachers and the researcher an advantage in the complex setting of piloting the tasks.

At the trial study most of the schools did not have a clear enough picture of the demands of the study. Even though they had received an information letter outlining the characteristics of the tasks and the necessary procedures, the most common preconception of the study was of students filling in an online questionnaire. To avoid problems of this kind in future it would be advisable to develop more detailed guidelines for conducting the trials at schools, including explanations of the implications for teachers and students regarding the time required. It would also be helpful to the schools to have the timetables well in advance so they can organize their own schedules.

Traditionally in Finland, schools and teachers have been eager to participate in the research activities. The Finnish model of teacher education has a strong focus on research-based understanding. However, during recent years the number of research projects has greatly increased and many schools have felt overloaded with contacts from researchers. This has made it more difficult to get schools interested in participation. However, the more collaborative the research activities are, the more willing the schools and teachers are to participate. The most important added value for Finnish schools and teachers comes from collaborative research projects which provide insights and tools for the development of pedagogical work. The various phases of the ATC21S task development clearly have the potential to provide this kind of benefit, although further thinking and development may be required to ensure that the sustainable effects of participation remain at the schools when the research is complete.

Conclusions

The introductory process of the assessment tools for 21st century skills has been an interesting and rewarding journey. All the phases of the study have been necessary and it would be very difficult to try to implement such a holistic assessment system without being part of the development process. The path from concept checking, via cognitive laboratories and pilots, to study trials has been fascinating and faced many obstacles, but has also provided a lot of knowledge, both about the tasks and about the ability of schools to adapt the assessment tool to their systems. The process has also introduced us to colleagues across the world. The international collaboration in different continents has itself proved the importance of fluency with various collaborative tools, but also revealed the necessity of solid 'good old tools' such as e-mail.

The project familiarized us in a detailed way with the theoretical background of the assessment, teaching and learning of 21st century skills. This has provided a good base for further utilization and development on the learning and teaching of these skills in Finland. The reactions from the students are promising but also show how little prepared they still are for collaboration and learning on the web. As Wells and Claxton (2002) point out, the socio-cultural element of learning no longer simply concerns transformation of a skill or knowledge; it now includes development of an understanding of the construction of mind and identity. The most important outcome of study lessons is what students are actually able to do with the skills they have learned (Silva 2009). Developing assessment tools can be seen as a very important step towards making them a more prominent feature of our everyday schooling. There is a demanding task ahead to bring the understanding of learning and assessment of 21st century skills through the different levels of our school system, from policy making to teacher training and classroom practice. If these assessment materials can convince Finnish teachers about their usability in teaching and learning the 21st century skills, a big step has been taken.

References

Arffman, I. (2012). Translating international achievement tests: Translators' view. Finnish Institute for Educational Research, Reports 44. Jyväskylä: Finnish Institute for Educational Research. Available at: http://ktl.jyu.fi/img/portal/22708/g044.pdf

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). New York: Springer.

Chesterman, A. (1997). *Memes of translation*. Amsterdam: Benjamins.

Grisay, A., Gonzalez, E., & Monseur, C. (2009). Equivalence of item difficulties across national versions in PIRLS and PISA reading assessments. In: *IERI monograph series:*

Issues and methodologies in large-scale assessments (Vol. 2, 63–84). Princeton: IERI-ETS Research Institute. Retrieved from: http://www.ierinstitute.org/fileadmin/Documents/IERI_Monograph/IERI_Monograph_Volume_02.pdf. Read 15 Feb 2014.

Holappa, A.-S. (2007). *Perusopetuksen opetussuunnitelma 2000-luvulla – uudistus paikallisina prosesseina kahdessa kaupungissa.* [The curriculum change as local processes in two cities.] Oulun yliopisto. Acta Universitatis Ouluensis E94.

Kankaanranta, M., & Puhakka, E. (2008). *Kohti innovatiivista tietotekniikan opetuskäyttöä. Kansainvälisen SITES 2006 -tutkimuksen tuloksia.* [Results from the international SITES 2006 -study]. Jyväskylä: Jyväskylän yliopisto. Koulutuksen tutkimuslaitos.

Kartovaara, E. (2009). *Opetuksen järjestäjien ja rehtoreiden näkemyksiä ja kokemuksia perusopetuksen vuoden 2004 opetussuunnitelmauudistuksesta*.[Feedback from the year 2004 curriculum change]. Helsinki: Opetushallitus.

Krokfors, L., Kangas, M., Vitikka, E., & Mylläri, J. (2010). Näkökulmia tulevaisuuden koulupedagogiikkaan. [Insights towards the school pedagogics of the future]. In R. Smeds, L. Krokfors, H. Ruokamo, & A. Staffans (Eds.), *InnoSchool – välittävä koulu. Oppimisen verkostot, ympäristöt ja pedagogiikka* (pp. 51–85). Espoo: Aalto yliopisto.

Lipponen, L. (2012). Ihmisenä kasvaminen [Human growth]. In E. K. Niemi (Ed.), *Aihekokonaisuuksien tavoitteiden toteutumisen seuranta-arviointi 2010* [Evaluation of the fulfillment of cross-curricular themes in 2010] (Koulutuksen seurantaraportit no. 1, pp. 43–58). Helsinki: Opetushallitus.

Ministry of Education. (2009). *Composite news bulletin*. June 2009. Retrieved from: http://www.minedu.fi/OPM/Tiedotteet/composite_news_bulletins/2009/June_2009.html?lang=fi&g in Finalndextra_locale=en. Read 25 Sept 2012.

Ministry of Education. (2012). *Tulevaisuuden perusopetus – valtakunnalliset tavoitteet ja tuntijako.* [Finnish Ministry of Education. The basic education in the future. Objectives and curricular basis]. Opetus- ja kulttuuriministeriön työryhmämuistioita ja selvityksiä 2012. p. 6.

Niemi, E. K. (2012). Opettajakysely [Teacher survey]. In E. K. Niemi (Ed.), *Aihekokonaisuuksien tavoitteiden toteutumisen seuranta-arviointi 2010* [Evaluation of the fulfillment of cross-curricular themes in 2010] (Koulutuksen seurantaraportit no. 1, pp. 19–39). Helsinki: Opetushallitus.

Norrena, J., Kankaanranta, M., & Ahonen, A. (2012). *Innovative teaching in Finland.* Paper presented at the 2012 American Educational Research Association conference, Vancouver.

Pohjola, K. (2011). *Uusi koulu. Oppiminen mediakulttuurin aikakaudella*. [New School. Learning the media cultural era]. Jyväskylä: Jyväskylän yliopisto, Koulutuksen tutkimuslaitos.

Salo, M., Kankaanranta, M., Vähähyyppä, K., & Viik-Kajander, M. (2011). Tulevaisuuden taidot ja osaaminen. Asiantuntijoiden näkemyksiä vuonna 2020 tarvittavasta osaamisesta. [Experts anticipating the skills needed in 2020]. In M. Kankaanranta & S. Vahtivuori-Hänninen (toim.), *Opetusteknologia koulun arjessa II* (pp. 19–40). Jyväskylä: Jyväskylän yliopisto, Koulutuksen tutkimuslaitos.

Siekkinen, T., & Saastamoinen, M. (2010). Esi- ja perusopetussuunnitelmien perusteiden 2000 ja 2004 toimivuus. [The functionality of the national core curriculum in preschool and basic education]. In E. Korkeakoski & T. Siekkinen (Eds.), *Esi- ja perusopetusjärjestelmän toimivuus.* Puheenvuoroja sekä arviointi- ja tutkimustuloksia (pp. 47–70). Jyväskylä: Koulutuksen arviointineuvosto.

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

Vähähyyppä, K., & Mikama, A. (2010). *Koulu 3.0.* [School 3.0]. Helsinki: Opetushallitus.

Välijärvi, J. (2011). Tulevaisuuden koulu vai kouluton tulevaisuus? [School of the future or future without school?]. In K. Pohjola (Ed.), *Uusi koulu: Oppiminen mediakulttuurin aikakaudella* (pp. 19–31). Jyväskylä: Jyväskylän yliopisto, Koulutuksen tutkimuslaitos.

Wells, G., & Claxton, G. (2002). Education for the learning age: A sociocultural approach to learning to learn. In G. Wells & G. Claxton (Eds.)*, Learning for life in the 21st century: Sociocultural perspectives on the future of education* (pp. 1–17). Oxford, UK: Blackwell.

Chapter 11
Case Study on the Implementation of ATC21S in the United States

Kathleen Comfort

Abstract Over the last decade the national dialogue in the U.S.A. around 21st century skills has reached critical mass in national competitiveness, workforce development and, most notably, in K-12 education circles (IMLS, Museums, libraries, and 21st century skills (IMLS-2009- NAI-01). Washington, DC, 2009). A prominent goal of education in the U.S.A. is to ensure that all students graduate from high school with the knowledge and skills necessary for life, college, and careers in a 21st century global economy (NRC, Education for life and work: developing transferable knowledge and skills in the 21st century. The National Academies Press, Washington, DC, 2012a). In order to reach that goal, several reform efforts have been underway nationally to provide the tools and resources necessary for teaching and assessing 21st century skills, and to change the standards for educating students from kindergarten through the end of high school.

Context

The Partnership for 21st Century Skills (P21) in conjunction with states and business partners has designed and implemented tools, resources, and a framework for 21st century learning in eighteen states (P21 2009a). The National Governors Association and the Council for Chief State School Officers developed Common Core State Standards in English language arts (NGA 2010a) and mathematics (NGA 2010b) that have been adopted by 44 states, the District of Columbia, four territories, and the Department of Defense Education Activity. The Common Core State Standards are aligned to 21st century skills in the P21 Common Core Toolkit (P21 2011).

In sync with Common Core efforts, the U.S. Department of Education's Race to the Top initiative supports consortia of states through two large testing projects – Smarter Balance Assessment Consortium (SMAC) and the Partnership for

K. Comfort (✉)
WESTED, San Francisco, CA, USA
e-mail: kcomfor@wested.org

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_11

Assessment of Readiness for College and Careers (PARCC) – to design assessments which measure the skills advocated by the Common Core State Standards. These new assessments move beyond standardized multiple-choice tests to include more innovative performance and technology-enhanced tasks, requiring students to research and analyze information, consider evidence, and solve problems relevant to the real world.

In science education, the National Research Council (2012b) developed *The Framework for K-12 Science Education*, the foundation for the newly released *Next Generation Science Standards* (Achieve 2013). The NGSS are aligned to the Common Core standards with correlations to 21st century skills. Other national efforts include the National Assessment of Educational Progress (NAEP) Technology and Engineering Literacy Assessment (TEL). The TEL assessment will focus on technology and engineering literacy as critical components of 21st century life (NAGP 2014).

National professional educational organizations, such as the National Science Teachers Association and the National Council of Teachers of Mathematics, have worked with the Partnership for 21st century skills to produce skill maps illustrating the integration of science and/or mathematics content with 21st century skills to influence curriculum and instruction (P21 2009b). The Institute of Museum and Library Services is working with museums and public libraries across the country to support them in envisioning and defining their roles within their local communities as institutions of learning in the 21st century and to enhance understanding among policymakers and other stakeholders about the integral roles museums and libraries play in creating an engaged citizenry and competitive workforce (IMLS 2009).

These national efforts seek to ensure that students will graduate from high school with skills that will make them college and career ready, as well as to ensure that they acquire and show progress in 21st century skills (NRC 2010).

ATC21S Development in the U.S.

WestEd implemented all ATC21S™project[1] activities in the U.S. WestEd is a preeminent educational research, development, and service organization with 600 employees and 16 offices nationwide. WestEd has been a leader in moving research into practice by conducting research and development programs, projects, and evaluations; by providing training and technical assistance; and by working with policymakers and practitioners at national, state and local levels to carry out large-scale school improvement and innovative change efforts. The agency's mission is to promote excellence, achieve equity, and improve learning for children, youth, and adults. In developing and applying the best available resources toward these goals,

[1] The acronym ATC21S™ has been globally trademarked. For purposes of simplicity the acronym is presented throughout the chapter as ATC21S.

WestEd has built solid working relationships with education and community organizations at all levels, playing key roles in facilitating the efforts of others and in initiating improvement ventures.

WestEd is recognized as a national leader in standards and assessment. WestEd works closely with schools, districts, and policymakers to ensure that all students – including English learners, the nation's fastest growing student population – meet the high learning expectations built into the new Common Core State Standards and the Next Generation Science Standards. WestEd also serves as the project management partner for the multi-state Smarter Balanced Consortium.

WestEd's STEM Program, spanning grades Pre-K through 16, contains a diverse portfolio of projects that enhance teaching and learning across all STEM subjects. Projects within the STEM program address cutting-edge research, evaluation, curriculum development, assessment development and professional development. STEM staff work nationally and internationally to increase understanding of issues such as technology literacy, the use of simulations to enhance student learning and assessment, and science and mathematics learning.

STEM staff have participated in the ATC21S project from its inception. STEM staff, recognized nationally and internationally for their expertise in 21st century skills and assessment, contributed to two ATC21S white papers. STEM staff also served as the National Project Manager for the U.S. by: representing the U.S. at international meetings; co-developing products with researchers in other countries; and implementing all phases of the project.

The work of ATC21S has significantly informed the work of STEM in defining and developing assessments of 21st century skills. WestEd STEM staff have been working with researchers at Stevens Institute of Technology in Hoboken, N.J. for the last 3 years on two National Science Foundation projects to develop and score paper-based assessments of 21st century skills at the high school and middle school levels. At the high school level, for example, students engaged in a collaborative problem-solving task to solve a real world problem about sequestering carbon on land and in the ocean. The task consists of an activity known as a "jigsaw." Four students worked together in a team and each student received a common piece of information describing the overall problem, and a unique piece of information critical to solving the problem. Students discussed the problem and shared their ideas and unique pieces of information. Each student was assessed on his or her ability to communicate the information provided by the other team members, as well as the reasoning behind the group's final solution to the problem. The results provided a measure of the extent to which the students were able to collaborate and communicate information during group problem solving (Sneider et al. 2012).

The comprehensiveness of the ATC21S framework helped STEM researchers to define the constructs measured in the Stevens assessments. While the ATC21S tasks are online and Stevens' tasks are paper-based, the coding procedures and collaborating abilities defined by ATC21S informed the development of rubrics and the scoring of the Stevens tasks. ATC21S also informed the development of assessments of 21st century skills for the Philadelphia Youth Network (PYN). The PYN 21st century skill set focused on both cognitive skills (e.g., communication, collaboration and teamwork, critical thinking) and on non-cognitive skills (e.g., initiative and self direction, productivity and accountability, flexibility and adaptability). ATC21S elaboration on the knowledge, skills, attitudes, values and ethics for each of the ATC21S skills in the framework was extremely helpful in defining the PYN 21st century skill set and in distilling the constructs for the PYN measures. The aim of the Philadelphia Youth Network is to measure 21st century skill gain in youth engaged in WorkReady Philadelphia programs. Other organizations in the city, such as the Philadelphia Academy, strive to prepare all Philadelphians to work and compete in the 21st century economy through its Digital On-Ramps initiative.

Method

Task Concept Review

The first phase of task development addressed the face validity of the task concepts for three Learning through Digital Networks – Information Communication Technology (ICT) literacy tasks and eight Collaborative Problem Solving (CPS) tasks. These tasks were intentionally designed for students ages 11 (grade 6), 13 (grade 8) and 15 (grade 10). While the LDN-ICT tasks – *Webspiration (Poetry)*, *Arctic Trek* and *Second Language Chat* – are set in scenarios that relate to typical content areas such as English language arts, science, mathematics, and languages, the constructs measured are skills within the LDN-ICT literacy domain. These skills include using a computer interface and search engines, posting questions and artifacts, collaborating with teammates, making tags, and performing basic IT tasks. The CPS tasks – *Balance, Plant Growth, Game of 20, Warehouse, Hexagons, Small Pyramids, Lightbox, and Sunflower* – are also set within typical subject matter content, e.g., mathematical and science contexts, and require students to communicate through an online chat box. The CPS tasks measure constructs in: (1) social skills, including participation, perspective taking, and social regulation; and (2) cognitive skills, including task regulation and knowledge building.

In implementing Phase 1, WestEd recruited 18 San Francisco Bay area teachers – six teachers each for grades 6, 8, and 10 – to preview the initial assessment prototypes and provide feedback on the task concepts as they related to their grade level. Nine teachers were assigned to review the LDN-ICT tasks and the other nine teachers were assigned to review the CPS tasks. WestEd staff worked with the ATC21S international research coordinator to develop training materials and protocols for collecting teacher feedback. In October 2010, WestEd held six training sessions – one for each age/grade level in LDN-ICT and one for each grade level in CPS. During each training session teachers were presented with an overview of the ATC21S project, introduced to task concepts, and worked with a partner to engage in and review the tasks on a live testing site.

For the LDN-ICT teacher training session, WestEd collaborated with the developers of the LDN-ICT tasks at the University of California, Berkeley (UCB). The primary aims of the task checking were to: (1) ascertain from the teachers whether they thought the tasks looked like they would be accessible to students; and (2) to find out whether the tasks had face validity for teachers. Secondary goals were: (1) to elicit from the teachers suggestions for modifications, deletions, or enhancements that would improve accessibility for students; and (2) to elicit from teachers how they thought the tasks were measuring the LDN-ICT construct.

Before previewing the tasks, and to better understand what the constructs in the tasks were measuring, the U.S. National Project Manager (NPM) provided an in-depth overview of the LDN-ICT construct. Within the area of LDN-ICT literacy, ATC21S focused on social networks, due to their function in contributing to learning to learn on the basis of enabling skills. The LDN-ICT construct is considered across four major functional areas: (1) functioning as a consumer – obtaining, managing and utilizing information/knowledge from shared digital resources and experts; (2) functioning as a producer – creating, developing and organizing information/knowledge in order to contribute to shared digital resources; (3) developing and sustaining social capital – using, developing, moderating, leading and brokering connectivity within and between social groups; and (4) developing and sustaining intellectual capital – understanding how tools, media and social networks operate and using these tools to build collective intelligence.

Next the NPM provided an overview of the three tasks and had the teachers log onto the LDN-ICT task website. The NPM walked the teachers through each task and described the tools built into the task that students would use. Following the overview of each task, teachers were asked to work with a partner and respond to a set of questions specific to the task. In *Webspiration*, teachers were asked: (1) What skills or capabilities do you think the tasks in the scenario are targeting? and (2) Considering the capabilities of your students, are there any questions or activities that should be eliminated from the scenario? For *Arctic Trek*, teachers were asked to: (1) Identify and write down three clues to retain and three clues to eliminate in the task; and (2) Discuss their rationale for retaining or eliminating the clues. For *Second Language Chat*, teachers were asked: (1) At what age do you believe native speakers would be able to learn and use a rating system? (2) At what age would native speakers be able to facilitate a chat topic? and (3) Suggest a chat topic that has the potential to engage language learners at the selected age.

At the end of the review session, the NPM asked the teachers to respond to five questions. A summary of teacher responses to the questions is shown in Table 11.1.

Teacher training for the CPS tasks was similar to the LDN-ICT training. The NPM provided an overview of the ATC21S project and of the CPS tasks and their constructs. The teachers were introduced to the Collaborative Problem Solving Framework, which described the aspects, contributions, and actions of the five constructs – participation, perspective taking, social regulation, task regulation and knowledge building.

The teachers logged onto the CPS website and worked with a partner to review five tasks. The teachers were provided with a set of questions to discuss as they

Table 11.1 Protocol questions and summary of teacher responses for the validation of the LDN-ICT task concepts

Protocol questions Does the task appear to:	Summary of teacher responses
Engage the students?	Engagement level is good, but 6th and 8th graders will need some front-loading to become familiar with the tasks and it may take them longer to complete. Students would enjoy the tasks.
Be appropriate for the students in terms of prerequisite knowledge?	Not all students have prior knowledge of computers. Students need a tutorial before beginning the task – even some teachers would be unfamiliar with the navigation tools. Students are introduced to concept maps by 9th grade and in algebra. The contexts of the tasks are relevant to students' lives.
Be appropriate for the students in terms of the socio-cultural context?	Yes, appropriate. However, English learners and students not familiar with standard English may be challenged.
Take a similar amount of time for students?	No. Timing depends on student prior technological knowledge and exposure to technology. English language learners and student with disabilities would need more time.
Differentiate between students?	Yes. Eighth graders will have a basic knowledge of web-based research and sixth graders are comfortable using chat tools. However, students who don't have computers at home and English learners may have trouble navigating between windows.

reviewed the tasks and their discussions were recorded. The teachers were also provided a form to record their comments for each task they reviewed. The form contained three questions and the first question contained a series of skills to be rated for each of the CPS constructs (Fig. 11.1). (Note that the actual form used brought together the Task Regulation and Knowledge Building strands of problem solving). The second two questions were: "Considering the capabilities of your students, are there any questions or activities that seem too far above or too far below the general range of capabilities?"; "What other comments do you have about the task?".

At the end of the review session, the NPM asked the teachers to respond to five questions. A summary of teacher responses to the questions is shown below in Table 11.2. All teachers were very enthusiastic about the opportunity to preview the new assessments and provided both oral and written feedback through small group discussions during the training sessions and through written responses in protocols. The protocols were designed to elicit teacher judgment of the relevance and appropriateness of each task for students of different ability levels in the grade level they were teaching.

Overall, teachers reported that both the LDN-ICT and CPS tasks were engaging for students and that students did not need to rely on prior content knowledge (e.g., English language arts, mathematics, science) to complete the tasks. Almost all teachers reported that the tasks were appropriate for students in terms of socio-cultural context, but cautioned that English learners and students unfamiliar with

Your Name ________________
Grade level you teach: ________

Item Name: ________________
Room Number: ______________

1. What skills or capabilities do you think the tasks in this item are targeting?

Participation

	Yes	No
Activity within environment	☐	☐
Interacting with, prompting, and responding to the contributions of others	☐	☐
Undertaking and completing a task or part of a task individually	☐	☐

Perspective taking

	Yes	No
Ignoring, accepting or adapting contributions of others	☐	☐
Awareness of how to adapt behavior to increase suitability for others	☐	☐

Social regulation

	Yes	No
Achieving a resolution or reaching a compromise	☐	☐
Recognizing own strengths and weaknesses	☐	☐
Recognizing strengths and weaknesses of others	☐	☐
Assuming responsibility for ensuring aspects of task are completed by the group	☐	☐

Problem solving

	Yes	No
Setting a clear goal for the task	☐	☐
Managing people or resources to complete a task	☐	☐
Formulating a course of action to address a problem or task	☐	☐
Implementing possible solutions to a problem and monitoring progress	☐	☐
Analyzing and defining a problem in familiar language (i.e. making the problem more manageable and meaningful)	☐	☐
Identifying need for further information	☐	☐
Developing lines of argument and explaining ideas to others	☐	☐
Accepting ambiguous situations and exploring options within these	☐	☐
Changing from one line of reasoning or course of action to another as information or circumstances change	☐	☐
Making connections between elements of knowledge	☐	☐
Follow path to gain knowledge	☐	☐

Fig. 11.1 CPS rating form and questions

standard English may be challenged. When asked about the timing of the tasks for all students, almost all teachers responded that younger students, English learners, and students with disabilities would require more time to complete the tasks. In terms of the tasks differentiating between students, almost all teachers responded that they thought the tasks would differentiate between students, but also raised equity issues regarding the availability of computers in the home. They cautioned that students without home computers might be at a disadvantage in completing the ATC21S tasks.

Table 11.2 Protocol questions and summary of teacher responses for the validation of the CPS task concepts

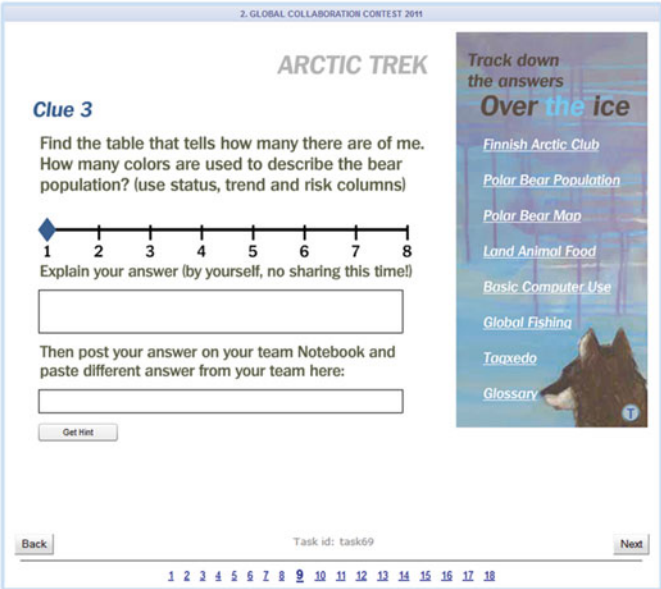
Protocol questions Does the task appear to:	Summary of teacher responses
Engage the students?	Any computer game would engage students. Students would be eager to solve puzzles, problems, compete against the computer, or collaborate with others if tasks are presented in a game format. But they need directions and clear objectives for them to overcome initial frustrations.
Be appropriate for the students in terms of prerequisite knowledge?	Some math knowledge can be useful in Hexagon and Small Pyramids. The ability to recognize patterns is useful. Creativity and spatial acuity helps. Otherwise, no subject-matter knowledge is required.
Be appropriate for the students in terms of the socio-cultural context?	Yes. Most students of any background are adept with texting, IM, and other forms of online chatting in English or a native language.
Take a similar amount of time for students?	No, 6th graders or 11-year-old students may need more time figuring out what's expected of them. They need more guidance. Some might give up too soon out of frustration.
Differentiate between students?	Yes. The assessments might be better served by having social or civic issue content, where students can collaborate on solving a social or community problem.

Cognitive Laboratory

The second phase of the project focused on the implementation of cognitive laboratories (coglabs) or think-aloud sessions with students. WestEd staff collaborated with the BEAR Centre at the University of California, Berkeley (UCB), to conduct sessions with 21 students representing grades 6, 8 and 10 in the San Francisco Bay Area in March 2011. The composition of the cognitive laboratory participants included: 17 % Asian, 22 % African American, 48 % Hispanic/Latino, 10 % White, and 1 % Multiple Ethnicity.

The think-aloud sessions were designed to gain insight into students' thoughts and perceptions as they navigated through the LDN-ICT tasks. In preparation for the coglabs, all student computers were checked to ensure they met technical requirements, including access to external websites. Berio screen capturing software was downloaded on all computers to "videotape" students as they worked through the tasks. Each student's login, talk-aloud and screen activity – e.g., pointing, clicking, using mapping tools, adding ideas, dragging and dropping movable items, accessing podcasts, on their desktop was recorded in real time.

Researchers observed students as they worked through the tasks ensuring that students were entering responses and talking simultaneously. As the researchers observed students, they also completed a cognitive laboratory template designed to record student screen actions. Figure 11.2 shows a sample protocol for the Arctic Trek task the researchers used to monitor student navigation through each screen of



Tasks:
Clue 3
❑ Use slider to indicate polar bear population.
❑ On 1st box, post individual answer.
❑ Cut and past answer from team notebook and enter on 2nd box.

Skills Applied			
ICT Consumer	**Y** ☐	**N** ☐	**N/A** ☐
ICT Producer	**Y** ☐	**N** ☐	**N/A** ☐
ICT Social Capital	**Y** ☐	**N** ☐	**N/A** ☐
ICT Intellectual Capital	**Y** ☐	**N** ☐	**N/A** ☐

Teach Aid:	**Y** ☐	**N** ☐
Usability Issues:	**Y** ☐	**N** ☐

Describe Issues:
Comments

ATC21S 10

Fig. 11.2 Protocol to monitor student navigation

each task. After completing the think-aloud session, the researchers interviewed students about their experience.

Since the LDN-ICT tasks are designed to measure digital literacy skills, students are discouraged from asking their teacher how to navigate through the tasks. The LDN-ICT tasks include a built-in feature – an "Ask three, then me" rule where students are expected to explore three information sources before asking the teacher for help. Students are expected to: (1) review the task directions and resources on their computer screen; (2) request and receive help from team members; and (3) access the Internet for information prior to asking for help. If a student requests teacher help, s/he must click on a teach aid button. Afterwards, the teacher records the type of help provided. During the cognitive laboratories, only a small number of students at grade 6 used the teach aid button to request help; students at grade 8 did not use it.

Overall, the cognitive laboratories generated valuable information about how well the LDN-ICT tasks worked, and which task features might need to be improved to make them work better. The *Webspiration* task was administered to students in all three grade levels. Researchers observed that most students had little difficulty logging on to the testing site, while only a few experienced usability issues, such as difficulty in uploading video. A few 6th grade students had problems finding a poem online, using radio buttons and the recycling bin, or adding and connecting ideas. At 8th grade, a small number of students experienced difficulty with entering text in a box, copying and pasting poems, and/or reordering questions. Students in the 10th grade did not appear to experience problems working through the activities. *Arctic Trek* was also administered to students in all three grade levels. Only students in 6th grade were observed to have difficulty with recognizing team members, finding the secret code, and tabbing between the Google Docs and the task page. During *Second Language Chat,* only administered at grade 6, only a few students had difficulty in viewing the help podcasts. Across all three tasks, most students were observed to be engaged in the tasks and able to complete most actions within the time allocations.

Pilot

The third phase of the project focused on pilot testing the LDN-ICT literacy tasks with students and teachers in real time in the classroom. The intent of the pilot was to produce a data set for clarification of coding and scoring purposes; to identify the feasibility of administering web-based assessments in a school setting; and to test the data capture and measurement methods involved in interpreting the complex responses that these types of assessment generate.

The pilot was undertaken in the classrooms of three teachers, one teacher for each grade level (6th, 8th, and 10th grades) in three schools in Northern California in September 2011. The teachers were recruited through WestEd's network of schools in the San Francisco Bay area. The 6th and 10th grade teachers participated in either the task concept review or the cognitive laboratory and expressed interest in the pilot test. The 8th grade teacher worked with STEM staff for several years assisting with the development and scoring of science assessments. Approximately 113 students participated (24 students in grade 6; 40 students in grade 8; and 49 students in grade 10). Across the three grade levels, the participants included: 23 % Asian, 15 % African American, 46 % Hispanic/Latino, 7 % White, 8 % Filipino, and 1 % Multiple Ethnicity.

The three teachers were trained to administer the LDN-ICT tasks to their classes. Training was conducted online using a 1-h webinar so that WestEd staff could walk the teachers through the administration process. The teachers received a copy of the PowerPoint for training and the Pilot Test Administration Manual prior to the training session which included: an overview of the ATC21S Project; the role of the teacher as facilitator in administering the pilot assessments; the administration of the LDN-ICT tasks, including student logins and passwords; and the technical requirements, including accessibility testing of individual student computers. Accessibility testing was necessary to ensure that student computers were able, with sufficient upload and download speeds, to access external websites. WestEd staff spent about 10 min per computer while assisting the 6th grade teacher. The accessibility testing added 2–3 additional hours to the teachers' prep time. The eighth and tenth grade teachers conducted the accessibility testing independently.

The teachers reported that their students were very enthusiastic about the tasks and very engaged. Students reported that they liked the interactivity of the tasks, as well as the opportunity to work and chat with team members in real time. Many students reported that they found the podcasts helpful and that they preferred the online format of the test to paper and pencil tests.

The 6th grade teacher reported that her students experienced difficulty logging on to the testing site, and that the server at UCB was extremely slow to non-responsive. UCB programmers reset the server to allow for more mistyping without triggering a security block. Other technical issues experienced at grade 6 included students not being able to save their poem on the *Webspiration* site, and not being able to download the Kudo game in *Arctic Trek*. The 8th grade teacher also reported that most students experienced problems logging into *Webspiration* and, as a result, were unable to communicate with their team members. His students had more success with *Arctic Trek* and really enjoyed the task. The 10th grade teacher reported that some students had problems logging into the testing website and that the assigned passwords/logins did not always work. She also reported that students experienced major confusion about working in groups and in finding their teammates. In a few cases, students in one class were pre-assigned to work with students in the other class. These students were not able to collaborate because their partners were not online – they were in a different class. Despite the technological glitches in locating teammates, a 10th grade student reported: "My most successful collaboration was that we had to answer the question and post them and share. This is successful because I think that it help give us other views of the answer to the question or if it was right or wrong." Other students reported: "I think this is an attempt to measure how well you can work with others without verbal contact. How successfully you can complete work as a team."

Trials

The fourth phase of the project focused on field-testing the CPS tasks in real time in the classroom. The CPS trials were administered in January 2012 in five states – Arizona, California, Colorado, New Mexico and Texas. Over 80 teachers applied and 21 were selected to best denote geographic location and grade level. WestEd recruited teachers by emailing a flyer announcing the trials to colleagues connected to large networks of teachers in several states.

Seven teachers were selected for each grade level (6th, 8th, and 10th grades) in 17 schools. Approximately 812 students participated in the field test (335 students in grade 6; 173 students in grade 8; and 304 students in grade 10). Across the three grade levels, the composition of the field trials participants included: 5 % Asian, 9 % African American, 38 % Hispanic/Latino, 46 % White, and 1 % Asian American.

The 21 teachers were trained via Webinar to administer the CPS tasks to their classes. A Power Point training presentation and a practice task were recorded on You Tube videos so that teachers could review the information at any time. The training session covered:

- A brief overview of the ATC21S Project
- The role of the teacher in administering the CPS assessments
- The administration of the CPS tasks
- The student logins and passwords
- The technical requirements for student computers
- Practice logins and passwords for teachers so they could try the practice task before administering the assessments to their students
- Helpful hints for successfully administering the tasks

All teachers were instructed to read and follow the instructions in the CPS Technical Specifications document to check student computers before students participated in the field trials. Teachers were then taken through a series of steps to check accessibility to the CPS web site housed at the University of Melbourne.

Eleven collaborative problem solving (CPS) assessment tasks were field-tested. Four tasks were categorized under "Puzzles and Experiments," including: Hot Chocolate, Clowns, Olive Oil, and Shared Garden. Seven tasks were categorized under "Mathematical and Scientific," including: Sunflower, Plant Growth, Warehouse, Balance, Hexagons, Game of 20, and Small Pyramids. Students were not administered all 11 tasks. Instead, they were administered the tasks in small bundles with students completing two bundles with their partner. Administration of the tasks took approximately 50–60 min providing time for students to log in, complete the tasks and answer a short online survey.

The process used for the student logins was much more streamlined than the LDN-ICT task process. Once the teachers scheduled the date and time for administration and informed WestEd of the number of students to be tested, the University of Melbourne sent WestEd unique student logins and passwords for each teacher. Since the students were working with a partner to complete the tasks, each login consisted of a student identification number (student001) and a team code (exam001). A two-student team was assigned as Student A and Student B during the login process. The practice task – "Light Box" – provided teachers and students with an example of the key features of the collaborative problem solving assessment tasks. The teachers logged into the University website with a partner and chose either the Student A or B role for themselves.

The key features of the CPS tasks were:

1. A chat box for students to communicate with their partner, located on the bottom right hand corner of the screen. Students entered messages in the "input message" window and clicked "send" to send the chat messages to their partner. A record of the conversation appeared in the rectangular box above the input message window.
2. A task environment that contained different views for Student A and Student B and different task resources for each student.
3. Buttons that allowed students to move on to the next page or exit the task (players could not go back to a previous page in the task).

While the administration and student login process for the field trials was streamlined and easy for teachers and students to follow, technological challenges were still experienced across all field test sites. Almost all schools experienced problems with firewall settings. School site personnel were hesitant to bypass their proxy and use a direct IP address from their firewall. One IT director told us that to do so would put his school out of compliance with the federal Children's Internet Protection Act requirements. Despite the technical difficulties, all teachers reported that students were highly engaged in the tasks and enjoyed the sessions. Teachers reported:

- "The test went well, and the opportunity to observe it in action has been priceless. I am grateful for the experience. The students were engaged and I feel even this small test experience has had an impact on their learning year."
- "The common thread was they enjoyed the test and are eager to hear any feedback they might receive from the data. Moreover, I was struck by their curiosity for the potential use of this type of assessment in their future education."
- "At first it was a bit cumbersome getting the students to understand how to work together problem solving, but as they dialoged through the text messaging application their apprehension decreased and their engagement increased."
- "As I observed the students during the assessment they appeared genuinely engaged in the series of tasks through experimentation, trial-and-error, and the interactive partner communication. It was interesting to view the differences in problem solving skills between partners, as well the ebb-and-flow of frustration levels as the students attempted to communicate next move dialog with each other."

Other findings include the observations of an Intel researcher in a middle school classroom in Rio Rancho, NM. According to the researcher, 10 years of Intel Teach data show that the primary barrier to effective LDN-ICT integration in educational settings is access and infrastructure. And, that this issue remains as the most significant barrier to the integration of LDN-ICT associated with the online assessments in the U.S. to date (Price 2012).

Challenges

According to Trilling and Fadel (2012), in today's digital age, jobs migrate around the globe and land with highly skilled individuals. It is more important than ever to equip young people with 21st century skills and to ensure that they can apply these skills to real-world challenges (Trilling and Fadel 2012). A recent Gallup study of Americans aged 18–35 found that "those with high 21st century skill development are twice as likely to have higher work quality compared to those who had low 21st century skill development (Gallup 2013, p. 4)." However, current research is showing that students are not learning 21st century skills because teachers are not teaching them (Saavedra and Opfer 2012).

The dominant approach to compulsory education in much of the world is still the transmission model through which teachers transmit factual knowledge to students through lectures and textbooks (RAND 2012). Students have the opportunity to learn information through the transmission model, but typically do not have much practice applying the knowledge to new contexts, communicating it in complex ways, using it to solve problems, or using it as a platform to develop creativity (RAND 2012). Experts agree that this is not the most effective way to teach 21st century skills (Boix Mansilla and Jackson 2011; Schwartz and Fischer 2006; Tishman et al. 1993). A second obstacle to students' development of 21st century skills is that they do not learn them unless they are explicitly taught, and these skills are not typically taught in separate stand-alone courses. Most teachers do not have sufficient experience teaching 21st century skills to have developed the deep expertise needed to train others.

The Common Core State Standards and the Next Generation Science Standards demand major conceptual shifts in curriculum, instruction and assessment. These new standards were intentionally designed to address college and career readiness through the lens of 21st century skills. They require students to exhibit critical thinking, problem solving, and other 21st century skills to demonstrate their understanding of concepts and apply this understanding to real-world scenarios (CCSS date; NRC 2012). If students are going to learn the 21st century skills needed to thrive in a global economy, then teachers need to have opportunities to learn how to explicitly teach 21st century skills. Student learning of 21st century skills requires 21st century teaching (Adamson and Darling-Hammond 2015).

In order to begin to address the major paradigm shifts in the new standards, and provide teachers with opportunities to learn how to teach 21st century skills, next generation professional development is needed that:

- Models the explicit teaching of 21st century skills in different content areas
- Provides opportunities for teachers to practice the teaching of 21st century skills in their content areas with students
- Provides coaching for teachers as they learn to teach 21st century skills
- Provides opportunities for teachers to learn how to assess 21st century skills and to review student work

- Provides opportunities for teachers to learn how 21st century skills fit into their curriculum and instruction and how to evaluate resources for teaching 21st century skills
- Helps teachers learn how to discuss 21st century skills with parents and other stakeholders
- Provides communities of learners among teachers for sharing experiences, strategies and resources

According to Bellanca (2014), the professional development of teachers is one of the most important areas to address in today's educational system. She maintains that teachers are being bombarded with 1 or 2-day lectures on how to implement new reform mandates and there is no way of knowing if they are implementing what they are learning or if they have actually learned anything. She argues that deep change is needed in teacher professional development to advance student learning of 21st century skills and that 'band aids' or quick fixes will not work.

Research shows that high quality professional development can have a powerful effect on teacher skills and knowledge and on student learning and achievement if it is sustained over time, focuses on core content, and is embedded in the work of professional learning communities that support improvements in teachers' practice (Gulamhussein 2013; Darling-Hammond et al. 2009). Research also shows quality professional development is intensive, sustained, well defined, and strongly implemented (Garet et al. 2001; Guskey 2003); is based on a carefully constructed and empirically validated theory of teacher learning and change (Ball and Cohen 1999; Richardson and Placier 2001); addresses core content and pedagogy (Gulamhussein 2013; Weiss and Pasley 2009); and promotes effective curricula and instructional models based on a well defined and valid theory of action (Cohen et al. 2002; Hiebert and Grouws 2007; Rossi et al. 2004). Effective professional development addresses concrete challenges in teaching and learning, and specific content, rather than abstract educational principles or teaching methods taken out of context (Darling-Hammond et al. 2009). Teachers are more likely to try out practices that have been modeled for them (Gulamhussein 2013; Penuel et al. 2007; Snow-Renner and Lauer 2005; Desimone et al. 2002), and they judge learning opportunities more valuable when they are hands-on, build on their own content knowledge, and provide strategies for teaching the content to their students (Gulamhussein 2013).

Conclusion

Several efforts have been underway over the last decade in the U.S. to incorporate attention to 21st century skills in educational standards and to provide the tools and resources for teaching and assessing them. The Partnership for 21st Century Skills in conjunction with states and business partners has designed and implemented tools, resources and frameworks. New national Common

(continued)

Core State Standards and Next Generation Science Standards have bolstered the implementation of 21st century skills in English language arts, mathematics and science, grades K-12. National assessment consortia have built innovative computer-based items and performance tasks to address the assessment of critical thinking, communication and problem solving in conjunction with reading, writing and mathematics content in new state testing programs.

While there was great enthusiasm from teachers and students who participated in the implementation of the four phases of ATC21S, we are still finding that a primary barrier to effective LDN-ICT integration in educational settings in the U.S. is access and infrastructure (Price 2012). As a first step in preparing for computer-based testing, the national consortia have developed Technology Readiness Tools to support all states and districts to prepare for the launch of national computer-based testing. By beginning to address the technology issues that have plagued schools and districts for the last decade, the U.S. is beginning to consider possibilities and pathways for the adoption of ATC21S. According to Adamson and Darling-Hammond (2015), the integration of research and development with classroom practices for developing 21st century skills through ATC21S assessment tasks and tools is critical to guiding the implementation and scaling of ATC21S. WestEd has supported the integration of research and classroom practice leading to the adoption of ATC21S through concept papers and proposals to federal agencies.

References

Achieve. (2013). Next generation science standards. http://www.nextgenscience.org/nextgeneration-science-standards

Adamson, F., & Darling-Hammond, L. (2015). Policy pathways for twenty-first century skills. In P. Griffin & E. Care (Eds.), *The assessment and teaching of 21st century skills: methods and approach* (pp. 293–310). Dordrecht: Springer.

Ball, D. L., & Cohen, D. K. (1999). Developing practice, developing practitioners: Toward a practice-based theory of professional education. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 3–32). San Francisco: Jossey-Bass.

Bellanca, K. (2014) Professional learning experiences: there is no quick fix. Http://p2.org/news-events/p21blog/1422-professional-learning. 27 May, 1(4), No. 14.

Boix Mansilla, V., & Jackson, A. (2011). *Educating for global competence: Preparing our youth to engage*. New York: CCSSO, Asia Society.

Cohen, D. K., Raudenbush, S., & Ball, D. L. (2002). Resources, instruction, and research. In F. Mosteller & R. Boruch (Eds.), *Evidence matters: Randomized trials in education research* (pp. 80–119). Washington, DC: Brookings Institution Press.

Darling-Hammond, L., Chung Wei, R. A., Richardson, A. N., & Orphanos, S. (2009). *Professional learning in the learning profession: A status report on teacher development in the United States and abroad*. Oxford: National Staff Development Council.

Desimone, L. M., Porter, A. C., Garet, M. S., Yoon, K. S., & Birman, B. F. (2002). Effects of professional development on teachers' instruction: Results from a three-year longitudinal study. *Educational Evaluation and Policy Analysis, 24*, 81–112.

Gallup. (2013). 21st century skills and the workplace; A 2013 Microsoft Partners in Learning and Pearson Foundation Study. http://www.gallup.com/poll/162818/21st-century-skills-linked-work-success.aspx

Garet, M. A., Porter, A. C., Desimone, L., Birman, B., & Suk Yoon, K. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal, 38*(4), 915–945.

Gulamhussein, A. (2013). *Teaching the teachers: Effective professional development in an era of high stakes accountability*. Alexandria: Center for Public Education.

Guskey, T. R. (2003). What makes professional development effective? *Phi Delta Kappan, 84*(10), 748–750.

Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students' learning. In F. K. Lester (Ed.), *The second handbook of research in mathematics education*. Reston: New Age and National Council of Teachers of Mathematics.

Institute of Museum and Library Services. (2009). *Museums, libraries, and 21st century skills (IMLS-2009- NAI-01)*. Washington, DC: Institute of Museum and Library Services.

National Assessment Governing Board (NAGB). (2014). *Technology and engineering literacy framework for the 2014 National Assessment of Educational Progress*. Washington, DC: Author.

National Governors Association Center for Best Practices, & Council of Chief State School Officers. (2010a). *Common Core State Standards for English language arts and literacy in history/social studies, science, and technical subjects*. Washington, DC: Authors.

National Governors Association Center for Best Practices, & Council of Chief State School Officers. (2010b). *Common Core State Standards for mathematics*. Washington, DC: Authors.

National Research Council. (2010). *Exploring the intersection of science education and 21st century skills: A workshop summary*. Washington, DC: The National Academies Press.

National Research Council. (2012a). *Education for life and work: Developing transferable knowledge and skills in the 21st century*. Washington, DC: The National Academies Press.

National Research Council. (2012b). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: The National Academies Press.

Partnership for 21st Century Skills (P21). (2009a). *Framework for 21st century learning*. Tucson. http://www.p21.org/storage/documents/P21_Framework.pdf

Partnership for 21st Century Skills Maps (P21). (2009b). http://www.p21.org/storage/documents/21stcskillsmap_science.pdf. http://www.p21.org/storage/documents/P21_Math_Map.pdf. http://www.p21.org/storage/documents/21st_century_skills_english_map.pdf. http://www.p21.org/storage/documents/ss_map_11_12_08.pdf

Partnership for 21st Century Skills. (2011). *Common Core Toolkit: A guide to aligning the Common Core Standards with the framework for 21st century skills*. http://www.p21.org/storage/documents/P21CommonCoreToolkit.pdf

Penuel, W., Fishman, B., Yamaguchi, R., & Gallagher, L. (2007). What makes professional development effective? Strategies that foster curriculum implementation. *American Educational Research Journal, 44*(4), 921–958.

Price, J. (2012). Classroom Observation in Rio Rancho, NM.

RAND Corporation. (2012). In A. R. Saavedra & V. Darleen Opfer (Eds.), *Teaching and learning 21st century skills: Lessons from the learning sciences*. Santa Monica: RAND Corporation.

Richardson, V., & Placier, P. (2001). Teacher change. In V. Richardson (Ed.), *Handbook of research on teaching* (4th ed., pp. 905–94). Washington, DC: American Education Research Association.

Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A systematic approach* (7th ed.). Thousand Oaks: Sage.

Saavedra, A., & Opfer, D. (2012). Learning 21st—century skills requires 21st—century teaching. *Kappan, 94*(2), 8–13.

Schwartz, M. S., & Fischer, K. W. (2006). Useful metaphors for tackling problems in teaching and learning. *About Campus, 11*, 2–9. doi:10.1002/abc.154. Wiley.

Sneider, C., Brockway, D., & Comfort, K. (2012). *Performance assessment of collaboration and communication skills: a work in progress*. Paper presented at the second 2nd P-12 engineering and design education research summit. Washington, DC, 26–28 Apr 2012.

Snow-Renner, R., & Lauer, P. (2005). *Professional development analysis*. Denver: Mid-continent Research for Education and Learning.

Tishman, S., Jay, E., & Perkins, D. N. (1993). Teaching thinking dispositions: From transmission to enculturation. *Theory into Practice, 32*, 147–153.

Trilling, B., & Fadel, C. (2012). What talents do young people need to thrive in the 21st century? A game of skills. *Royal Society for the Encouragement of Arts, Manufactures and Commerce. Journal*, 22, 10–15.

Weiss, I. R., & Pasley, J. D. (2009). *Mathematics and science for a change: How to design, implement, and sustain high-quality professional development*. Portsmouth: Heinemann.

Chapter 12
The Adaptation and Contextualization of ATC21S™ by Costa Rica

Maria Eugenia Bujanda and Elsie Campos

Abstract One of the most important features of the ATC21S™ project has been the creation of a Latin American Chapter represented by Costa Rica (The acronym ATC21S™ has been globally trademarked. For purposes of simplicity the acronym is presented throughout the chapter as ATC21S.). At the request of the Inter-American Development Bank (Banco Interamericano de Desarrollo) and with support from Intel and Microsoft Latin America, the ATC21S consortium authorized the creation of a Latin American Chapter and, within it, Costa Rica as the first partner country. The project in Costa Rica is known as the Assessment of 21st Century Competencies, and was implemented by the Costa Rican Ministry of Public Education and the Omar Dengo Foundation, with local support from the Costa Rica-United States Foundation for Cooperation. The project was designed as a pilot for the introduction of the experience throughout Latin America, so it aimed generally to validate the tools developed by the global project for the measurement of 21st century skills and their contextualization in Latin American countries.

Context

Costa Rica is known to be a country that values education. Education was made constitutionally universal and free in 1847, at a much earlier date than most of the, at that time, more developed countries in the region. In 1987, the country took the decision to introduce computers in every primary school. In 2011, the Constitution was reformed so as to establish as mandatory for the State an investment in education of a minimum of 8 % of the Gross Domestic Product.

The Costa Rican education system has 940,000 students, 73,616 teachers, and 4,523 schools (MEP, 2014). It is managed nationally by the Ministry of Public

M.E. Bujanda (✉)
Omar Dengo Foundation, San Jose, Costa Rica
e-mail: maria.bujanda@fod.ac.cr

E. Campos
Ministry of Public Education, San Jose, Costa Rica

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_12

Education. The Ministry implements the policies and the curriculum approved by the National Board of Education (*Consejo Superior de Educación*), a high level constitutional body responsible for defining the country's educational policy and thus above the Ministry, although at the same time chaired by the Minister. The Ministry decides on the curricula and the projects that have national scope, as well as the strategies and guidelines that shape the national education system.

Regional offices, of which there are 27, help the Ministry to organize and supervise the delivery of the educational services through schools, and to contextualize the educational policies. Locally, the schools have the option of developing an annual working plan, on the basis of their specific strengths and needs, as part of a continuous improvement process.

The Costa Rican education system's structure comprises three major levels:

1. *Preschool* (2 years, 5–6 years of age)
2. *General Basic Education*:
   - Primary school (6 years, 6–12 years of age)
   - Lower Secondary school (3 years, 13–15 years of age) taught by teachers of different specialties
3. *Diversified Education* (2 or 3 years, depending on the branch; 16–18 years of age); the main branches are Academic (2 years) and Technical (3 years), with other branch options such as artistic, scientific, etc.

Since 1987, the Ministry and the Omar Dengo Foundation (ODF) have run together a national program that has led the introduction of digital technologies in the schools, first in the form of computer laboratories, now in combination with mobile technologies in the classroom (with several modalities: the one to one model in some cases, and mobile labs in others). In 2014 the National Program of Educational Informatics reaches 71 % of the student population. It plans to reach 100 % by 2017.

One particular feature of this program has been a *constructivist* vision of the potential of computers in education as tools to think and to create, and thus to promote higher order skills such as problem solving, creativity, collaboration, etc. This approach to technology in education has inspired a set of projects that have as their basis the learning of programming.

More recently, the Ministry of Education launched a series of curriculum reforms directed towards making the learning more relevant and attractive for the students and to develop social and personal competencies that have traditionally not been part of the education system's objectives. A particular aim for Costa Rica, through its active participation in the ATC21S project, was to strengthen the assessment component of these recent curriculum reforms implemented during recent years, in addition to providing Latin American countries with educational resources for assessing and teaching 21st century skills through the promotion of efficient use of the Information and Communications Technologies (ICTs). In this context, the

Costa Rican activities in adaptation of tasks provide a state-of-the-art model for countries wishing to modify these complex online tasks for local use. The following provides information on the adaptation process for the tasks.

ATC21S Development in Costa Rica

The Latin American Chapter of ATC21S was implemented in Costa Rica by the Ministry of Public Education and the Omar Dengo Foundation, with the support of the Inter-American Development Bank (IDB), Intel and Microsoft Latin America, and the Costa Rica-United States Foundation for Cooperation (CRUSA).

The Inter-American Development Bank played a major role in bringing Costa Rica into the project. Its argument about the value of having a Latin American country to help contextualize and validate the assessment instruments in the context of the developing world convinced the ATC21S consortium of the relevance of Costa Rica's participation in the project.

The Costa Rican Ministry of Education was also quick to accept the invitation to be part in ATC21S, with the technical support of ODF, as it saw the initiative as a great opportunity to understand, through new approaches to assessment, how students progress in their learning of these crucial skills and how to support them in this process. The IDB's proposal also obtained an enthusiastic response from other partners that were invited to join the initiative: Intel Latin America, Microsoft Latin America, and CRUSA Foundation.

The Ministry of Education assigned a National Project Manager to lead the country's participation and conducted the necessary arrangements with the IDB and the rest of the supporting partners to officially represent Costa Rica as a member of the international project. Other roles that the Ministry took on were to invite the participating schools to be part of the project and to provide the necessary approval for the teachers to participate in the training workshops. A group of curriculum specialists from the Ministry helped in the process of localizing some of the assessment tasks. In the latest phase of the project, it also took a leading role in the production of digital resources aimed at helping teachers and students explore new ways of teaching and learning key competencies.

The Omar Dengo Foundation designated a National Project Coordinator and took on the responsibility for the technical implementation and for the administrative and financial management the project. A team of three researchers, including the National Project Coordinator, conducted the process needed to localize the assessment system, as well as the field research activities and reporting to the University of Melbourne. Finally, the Omar Dengo Foundation contributed with the creation of dissemination materials, including a website, a digital interactive booklet, and a series of videos showing examples of best teaching practices identified among participating teachers.

Localisation

As a Spanish speaking country, one of the first needs that Costa Rica encountered when it became a member of ATC21S was to translate the assessment instruments into the local language. Unlike countries such as Singapore or even Finland, where higher levels of English literacy can be assumed, the Costa Rican student population, in general, does not have an English proficiency level to guarantee its valid participation in tests written in that language.[1] In addition, the role assigned to the country as representative of the Latin American region was to facilitate the development and implementation of a set of assessment tools translated into standard Spanish.

Beyond the translation of all tasks to standard Spanish, the localisation of two tasks developed to assess the skill Learning through Digital Networks (LDN), involved additional work. These two tasks, *The Arctic Trek* and *Webspiration (Poetry)*, both require students to explore web resources that are external to the tasks. This proved to be a major challenge for the team, due to the difficulty of finding Spanish language web resources similar enough to those used in the English version of the tasks. Therefore, the team produced from scratch a set of web resources in Spanish language to substitute the original ones linked to the tasks in English.

Normally, it would not be much of an issue for students to be able to access Spanish language sites for their school-based research, because of the growing body of good quality content in this language on the Web. But in this case, it was required that the resources would match as closely as possible those used in the original tasks, so as to not risk the equivalence of the level of difficulty of the items included in the tasks.

The Arctic Trek

The adaptation process for the *Arctic Trek* was divided into three stages:

1. The search for a theme for which online resources exist in Spanish, and which are equivalent to those used in the original task
2. Selection of appropriate web resources and modification of the text within the task, according to information for the new theme
3. Creation of web resources that could not be found online, as well as new design elements.

In the case of *Arctic Trek*, the major challenge was to find external web resources that were the equivalent of those used in the original English-language task, in large part high-quality scientific sources on the Arctic. Due to this limitation, it was necessary to identify a theme that was closer to the Latin American context. A panel composed of the research team, the National Project Manager and science advisors,

[1] In spite of this, it is worth mentioning that the first stages of the field activities (the first cognitive labs and pilots) were implemented before the assessment platform was fully translated into Spanish, and that this was possible due to the high level of English literacy of students from several bilingual public high schools that were invited to be part of the project.

considered as its first option the rain forest, a bioclimatic landscape present in an important part of Latin America. However, the panel ultimately chose to work with the Antarctic, another Latin American landscape that would tie in with the Arctic theme in the original task.

The next step was to locate online resources with information on the Antarctic that met the following requirements:

- Belong to recognized educational-scientific organizations that could be considered trustworthy sources of information; preferably not personal websites or blogs
- Discuss topics relating to the Antarctic
- Have maps, tables or numerical data that could be used in different parts of the task, as was done in the original task in English
- Be interesting and engaging for children and adolescents
- Demonstrate a level of complexity appropriate for each age group (including aspects of content and language)
- If possible, be ad-free.

In the third stage, the text for the task was modified to correspond to the new theme and the selected web resources – mainly the clues for each age group that would be included in the developed task. In the same way, some parts of the task that contained graphics had to be redesigned to match the updated theme. Following the example of the polar bears and Arctic wolves in the original task, the team compiled information on animal populations in the Antarctic – specifically, penguins – and this information was used to create the new graphics. For the final creation of the measurement-based graphics, an expert in mathematics was called upon to define the ranges that the lines should trace.

The final phase of adaptation consisted of the ad hoc creation of a Spanish-language web resource to substitute the "Tagxedo" website, a resource used as filler in the original task. Tagxedo is a web resource that enables creativity by building word clouds. Its role in the task is inclusion in the list of websites that are available to students and from which they can obtain the information needed to answer the questions that the task poses to them. Due to the lack of a similar resource in Spanish language, in its place a website was designed that would allow visitors to draw like the artist Jackson Pollock, with information on this famous American painter and on expressionism (visit http://atc21s.mep.go.cr/jackson-pollock/).

The contextualization work concluded with the creation of a new graphic design for the whole task. The original task in English language used a graphic design based on Arctic wolves which in the Spanish version were substituted with Antarctic penguins.

Webspiration (Poetry)

In the case of this task, the adaptation process consisted of four major steps:

1. The selection of three poems in Spanish to replace those used in the original English-language task

2. Modification of texts within the task, according to the new poems
3. Compilation of web resources to replace those used in the original task
4. Creation of web resources that could not be found online.

Adaptation began with the selection of Spanish-language poems to replace those used in the original English-language task. For this, Spanish literature experts were consulted, who made recommendations about themes, language and figures that were appropriate for the ages of the students who would be doing the tasks. Cognitive laboratories with students allowed the team to identify that one of the three poems originally selected presented too high a level of difficulty, and it was replaced. The final selection includes a poem by José Martí, a Cuban poet from the nineteenth century; another from Amado Nervo, a Mexican poet who wrote at the beginning of the twentieth century; and finally one from Raúl Aceves, a contemporary Mexican poet who kindly granted author rights to his poem. With the poems selected, the team proceeded to modify the text concerning the poems to mirror the logical rules employed and the desired difficulty level for each age group from the original English language based task.

The next stage in the adaptation process for this task consisted of the substitution of web resources used in the task by their equivalents in Spanish. It was possible to find existing resources or websites for:

- Poetry terminology
- Authors
- Dictionary
- Glossary

The criteria for their selection were the educational content of each and their relationship to a trustworthy and serious institution. For example, for the "Dictionary" resource, the online Dictionary of the *Real Academia de la Lengua Española* was chosen – the official organization in the Spanish speaking countries in charge of protecting and promoting the good use of the language.

Finally, as it was not possible to find an equivalent resource in Spanish for the website FavoritePoem.org, it was necessary to create an ad hoc resource. The consequent website in Spanish is called Our Favorite Poems (visit http://www.atc21s.mep.go.cr/nuestrospoemasfavoritos/) and presents 18 videos of people with different backgrounds reciting their favorite poems and explaining the significance the poems have in their lives.[2] In the production of this resource, particular attention was paid to representing the ethnic and cultural diversity that exists in Latin America. For example, the participants in the videos come from different countries of the region (Argentina, Chile, Costa Rica, Nicaragua, Perú, etc.), and the nationalities of the poets they discuss also vary. In addition, an Afro-American

[2] The "Our Favorite Poems" website is exceeding the scope of the project and has been included in activities that promote reading, such as those hosted by the Ministry of Education in 2012 during the Week of Books (Semana del Libro). For the Ministry of Education, there exists a clear relationship between 21st century skills, the promotion of reading comprehension and the love of reading in general.

person and an indigenous person participated, talking about poetry in their respective cultural backgrounds and reciting a poem representative of them.

Adapting these tasks into Spanish required consideration of issues that were both relevant and significant with regard to their application, such as cultural diversity, language, varying geographical scenes, and some limitation on the use of existing online resources. Throughout these adaptations, the primary goals for the Costa Rican team were for the assessment tools to be valid for the rest of the region, as well as responding to the curriculum reforms that are occurring in Costa Rica.

Method

Task Concept Checks, Cognitive Laboratories, Pilot Programs and Trials

In 2011, the team's work efforts were concentrated on the fieldwork for *Collaborative Problem Solving* (CPS) and the technical feedback for the international research team. In two high schools, the research team conducted concept checks with eleven Science, Mathematics and English teachers, and cognitive laboratories with twenty 15-year-old students.

Pilots were conducted in two elementary and two high schools, with four groups of 11-year-old students, two groups of 13-year-old students and two groups of 15-year-old students. Eight Science, Mathematics and Educational Informatics[3] teachers took part in a four-hour training session, where they were familiarized with the project, its objectives, the tasks and the administration guidelines. Then the teachers administered the tasks. The results of these activities were reported to the team at the University of Melbourne, along with suggestions for the task administration procedures to be used in Costa Rica.

In the trials sixteen elementary and high schools collaborated throughout the country, with the participation of 90 11-year-old students, 222 13-year-olds, and 188 15-year-olds. Forty teachers of various subjects, including Mathematics, Science, Spanish and Educational Informatics (and the principals of some of these schools) participated in a four-hour training session in preparation for the administration of the tasks.

In 2012, new trials for the CPS tasks were conducted, this time in eleven high schools, with 593 13 and 15-year-old students. A two-day workshop was organized with the participation of 30 teachers in order to allow them to learn not only about the tasks and how to administer them, but also about the 21st century skills movement, the Collaborative Problem Solving skill and the assessment rationale behind ATC21S. When the trials were concluded, the teachers shared their experience with the assessment tasks in a 2-day briefing workshop, providing insights about how to use these tasks and develop didactic interventions that could help the students develop the skills.

[3] Computation Science teachers.

In 2012, we conducted cognitive laboratories for the *Learning through Digital Networks* (LDN) tasks with twelve students (four 11-year-olds, four 13-year-olds and four 15-year-olds) in one elementary school and three high schools. This activity provided valuable information about students' reactions to the localised tasks.

In 2013, another set of trials was conducted, this time combining CPS tasks and, for the first time in Costa Rica, LDN tasks. The goal was to provide additional CPS data to complete the psychometric calibration of the tasks, and also to test the Spanish-language version LDN tasks technically – both in terms of task-specific bugs as well as issues associated with the assessment system and its platform including registration and reporting. Of the 776 13-year-old and 15-year-old students from nine high schools who participated in these new trials, 575 students provided full scorable data, the loss due in part to school connectivity issues and task non-completion.

The 24 teachers who took part in these 2013 trials had also participated in the project in 2012 and were well familiarized with the CPS tasks, but as this was not the case with the LDN tasks and new features of the assessment system, additional training was organized – again a two-day workshop prior to the task administration and a two-day workshop after it. In this last workshop, the teachers had the chance to review the students' CPS reports, compare them with what they would have expected of their students, and reflect on how to use the information provided by the reports to make decisions about didactic interventions with their students.

In every stage of the field work, the sample of students was opportunistic. The participating schools were public in all cases except for one. This means that the socio-economic background of the participating students can be assumed to be mainly middle and middle–low. The schools were selected on the basis of their technical infrastructure (computer labs in good condition and good connectivity levels) and the positive disposition of the principal and the teachers towards the project. These two conditions provided, the research team tried to ensure that the school sample was varied enough regarding their geographical localization (rural and urban) and their record of student performance.

Responses to ATC21S Tasks

The teachers' and students' responses to the tasks were generally very positive. Though there were some technical difficulties – mainly due to Internet connectivity in the schools – the students enjoyed the tasks and the teachers highly valued the opportunity to get to know and use such innovative tools to assess skills that, though regarded as very important, have not been part of their explicit teaching objectives.

The students found the tasks both engaging and challenging. In the words of a student at Naranjo Bilingual High School: "You have to think, it makes sense, and it is fun". They worked on them with persistence and interest, though at times they felt helpless and frustrated, especially those who faced technical difficulties. A frequent comment made by the teachers was how striking it had been for them to see their students concentrating so deeply for a period of over two hours – the average duration of the sessions. As one of the teachers at Palmares Bilingual High School put it: “It drove the students’ interest because it is a change of routine and because they like to be challenged” (Greivin Calderón, personal communication, 2011).

At the same time, something that was evident for the research team from classroom observations was that in general the students do not read thoroughly what is handed to them. Also, for many students, it was shocking, and even disturbing in some cases, not to be told precisely what to do. This may be due to the fact that they are accustomed to receiving complete and specific indications, so ATC21S tasks represent a big change from the way instructions are given in the daily life of Costa Rican classrooms. Another classroom observation of interest was the degree to which some students limited their exploration of the elements being displayed on their screen, as if they were unaccustomed to such lack of constraints, or had not had the opportunities to explore digital environments.

The teachers showed great interest in learning how to incorporate the assessment tasks into their subjects. They appreciated their value as a tool to assess the students’ skills and to inform their teaching practice, but also as learning resources in their own right. The Principal of Presbítero Manuel Bernardo Gómez Elementary School expressed “What I like of this assessment is that it allows students to explore, without being afraid to fail, something that they hardly experiment with in their classes” (Katerine Ramírez, personal Communication, 2011).

The schools managed the task administration sessions well, with some exceptions in which teachers did not follow guidelines precisely. These schools provided the research team with the opportunity to develop examples of how to do the breakout session with the students and how to engage them for the tasks using additional materials (such as an introductory presentation, for example) with background information about the project and the assessment. Some teachers found it difficult to restrain themselves from helping the students when the students experienced difficulties. As a result, teachers were encouraged to provide generic help to the students by giving advice such as:

- To read very carefully the instructions and identify vital pieces of information in them
- To consider the fact that partners may have different things on their screens and different resources
- To persist despite possible difficulties
- To trust the power of the collaboration among the team members to resolve the tasks
- To be focused on the work being done by the partner, not the person next to them, and to keep constant communication with the partner through the chat
- Not to leave the task before informing their partner and hopefully not before having reached an acceptable solution
- To keep silence.

Based on its experience, the Costa Rican research team recommended that:

- Schools need to have good internet connectivity (minimum of 4Mbps for a group of 18 students). In order to lessen connectivity burden, a good strategy is to have the student pairs do the task bundle in a different order, as this can avoid having all of them using the same resources at the same time.

- Teachers can make use of a presentation in order to introduce students to the goal of the assessment tasks and the importance of the skills being measured. This presentation needs to be written in a simple language, so that it can be easily understood by children. It can also make use of eye-catching images and motivational phrases so that students become engaged with the activity.
- Before starting the tasks with the computers, it is important to seat students around in a circle in order to explain the activity but mostly to motivate them, emphasizing the basic instructions the students should take into account.
- Students whose assigned student is 'A' must sit side by side and, if possible, away from students whose assigned student is 'B', because even though partners don't sit nearby, students tend to look at the screen of their neighbours so they can see what the complementary students have on their screens. If possible, a good idea is to have the pairs split up in different rooms.
- The teachers need to study well the administration procedure before initiating the process.
- It is important to verify that students know how to type their logins correctly and that they choose their assigned student ('A' or 'B') accordingly.
- Teachers must not tell students what to do, and they must not help them solve the tasks or answer the surveys. Students must discover what should be done and decide, together as a team, what they want to do, as well as when to move on to the next page or when to finish the task.
- Teachers can do a closing at the end, in a circle, so that students can comment about their experience with the tasks. They can be asked about what they liked the most, what they think these tasks assess, how they collaborated and what they think was the most difficult part. The learning gained through the process should be highlighted: the relevance of good communication skills, of knowing how to express their ideas in a clear way so that people understand them, of thinking with their partner about the problem and what is it about, and of trying different solutions and strategies, etc. One main thing to remind the students of is that these tasks are not measuring their intelligence but the strategies they use to solve problems, their persistence, their willingness to collaborate and their skill at it. This is important, as many students may see the assessment tools as very challenging and only achievable by the brightest.

Challenges

Strong Internet connectivity is vital for the effective use of the ATC21S system. Considering that the tasks were conducted with relatively small groups of students (18 on average) and that most of the participating schools had 4 Mbps connectivity, there are concerns in countries such as Costa Rica that there is not sufficient technical infrastructure capacity to support such initiatives at large scale.

Learning how to use technology effectively in their classrooms was something that the teachers considered to be vital for them to be able to teach and assess 21st century skills. Recognizing themselves as digital immigrants, their main concerns were, first, that they are falling behind their students in terms of knowledge of digital tools and digital practices and, second, that they do not have appropriate equipment and resources for incorporating digital technologies in the classroom. Despite this, they showed willingness to learn new things, asking for ongoing training and support.

Teachers expressed the need to better know and understand what each of the skills involves in terms of components and specific behaviours, as well as about the ways these can be learned. This is closely related to knowing how to translate this information into specific didactic interventions.

Beyond the incorporation of new didactic strategies and digital technologies in their teaching practice, Costa Rican teachers talked about the need to make more profound changes in the education system. They talked about the need for a shift of paradigm: to change our current approach to education as an individual action and begin to understand it as a collective and social action. They believe that this must be accompanied by classes that encourage autonomy and responsibility of students for their own learning.

Such a change requires a whole-school commitment. That's why the teachers recommended that the plans to generalize the use of ATC21S tools and approaches in our education system should include training opportunities available for every interested teacher, online resources to support their work, but also a school-centred professional development strategy oriented to promote collective efforts towards the development of these key skills for our children and young people.

In order to provide a starting point for these future developments, the ATC21S Latin American Chapter complemented its work in support of the localisation and validation of the assessment system with the following dissemination and production activities:

- A set of digital resources for students that promote 21st century skills, available at the portal of the Ministry of Education (visit http://www.mep.go.cr/educatico).
- A web site with information on the 21st century skills as defined by ATC21S and a resource bank with diverse teaching aids that support their teaching and assessment (visit http://www.fod.ac.cr/competencias21/)
- A collection of videos showing best practices as implemented by some of the Costa Rican teachers that participated in ATC21S (visit http://www.fod.ac.cr/competencias21/index.php/areas-de-recursos/videos).
- A digital interactive publication that helps teachers acquire a general and practical overview of the main features that characterize the methodologies for teaching and assessment of 21st century skills (visit http://www.fod.ac.cr/competencias21/index.php/areas-de-recursos/publicacion-digital#.U5NoYXKwaQk).

Additionally, the *International Forum ATC21S: Assessment and Teaching XXI Century Skills* was organized in San José, Costa Rica, on 2 and 3 April 2014 (visit http://www.fod.ac.cr/competencias21/index.php/areas-de-recursos/comunidad/foro-internacional-atc21s#.U5NkM3KwaQk). The goal was to create a space for analysing and reflecting on the teaching, learning and assessment of 21st century skills, linking with educational reforms promoting MEP. It was attended by international experts such as Claire Scoular, Lead Researcher at ATC21S from the University of Melbourne, Australia; Eugenio Severin, Chilean consultant in education and technology; Moritz Bilagher, regional coordinator for Evaluation and Analysis of Trends in the Regional Bureau of Education for Latin America and the Caribbean of UNESCO; and Alberto Cañas, associate director of the Institute for Human and Machine Cognition.

Conclusion

In addition to participating with other partner countries in the fieldwork aimed at validating the assessment tools, the Costa Rican team has achieved the rigorous translation, adaptation and, in one case, creation of these instruments. This component of the project has contributed to understanding how linguistic and cultural differences affect the assessment of skills and competencies in the 21st century.

Within the strategy of the Costa Rican Ministry of Education, participation in the ATC21S project was part of a program to promote the reform of teaching and assessment practices through the use of technology and the development of better teaching and learning tools and approaches. These efforts join a policy that aims to make the most of the use of mobile technologies, the educational informatics laboratories in schools and the application of innovative, more meaningful and participatory learning methodologies in the curriculum.

The set of skills defined by the ATC21S project closely relates to the programs of study developed within the Costa Rican curriculum under the "Ethics, aesthetics, and citizenship" program (Civic education, Fine arts, Education for life, etc.). They also coincide with an emphasis on key skills in the reform of subjects like Mathematics and Spanish. For example, into the Spanish program of study in Secondary Education there were introduced elements of logic, in order to develop the ability to detect reasoning, build arguments and engage in debates in different contexts, and to improve students' reading comprehension as well. In Primary Education, the "Think Art/Piensa en Arte" program was implemented in Spanish classes as a way to strengthen critical and rational thinking via the employment of art as a pedagogical resource.

Given the preceding conditions, the Assessment and Teaching of 21st Century Skills project helps to consolidate and advance important efforts to improve the quality and relevance of our education. The further work that still needs to be done will allow our education system to complete the process of migrating from a traditional to a more innovative system. That transformation will be a reality, thanks to what innovative teachers and committed school leaders, pushed by the strong impulse of their students, will build collaboratively to better respond to our society's most challenging needs and expectations.

Chapter 13
ATC21S Trials of Collaborative Problem Solving Tasks in the Netherlands

Diederik Schönau

Abstract The Netherlands joined the ATC21S™ project in April 2011, when the pilot studies were already underway in the other countries (The acronym ATC21S™ has been globally trademarked. For purposes of simplicity the acronym is presented throughout the chapter as ATC21S.). Due to this late entry into the project, the decision was made to skip the pilot phase in the Netherlands, as well as the concept checks and cognitive laboratories that had already been conducted in founder countries. It was decided to concentrate on the translation of the final products for use in the trials that were planned for November 2011. It was also decided to limit the trials to the Collaborative Problem Solving (CPS) tasks, because translation of the Learning through Digital Networks – Information Communications Technology tasks would generate complex problems beyond the resources of the local project.

Context

The Dutch basic national education system is divided into two stages: primary and secondary education. Primary education, running from age 4 to 11, has eight grades and is given in about 7,000 schools. After primary education, at age 12 years, students enter secondary education, which consists of three main levels: lower vocational education (with four pathways), which takes 4 years; senior secondary education, which takes 5 years; and pre-university education, which takes 6 years. There is no national curriculum. The only official guidelines are generally formulated attainment targets for primary education and more detailed examination programmes for all subjects at all levels in secondary education. Within those guidelines schools have the freedom and responsibility of giving form to their curriculum. In the 1980s and early 1990s of the twentieth century, fundamental reforms took place in both primary and secondary education, to adjust the education system to the major changes taking place in society. Greater importance was given to the so-called 'general skills' that would better prepare the students to take part in a more diverse civil

D. Schönau (✉)
CITO, Arnhem, The Netherlands
e-mail: Diederik.Schonau@cito.nl

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_13

society, to be in better command of their own learning skills ('learning to learn') and to better develop their own individual interests and skills. The growing influence of Information Technology was also taken into consideration, but since the mid-1990s technological developments have progressed enormously. In the Netherlands, it is left to schools to adjust their teaching and learning strategies to assimilate these developments into their daily practice. This has resulted in a great variety in the ways Dutch schools have organised their subjects, the time schedules and the way students learn, either as a group or individually, with teacher instruction or by computer assignments.

ATC21S Development in the Netherlands

The Netherlands became involved at a late stage in the ATC21S project, following a study visit of a Dutch educational delegation to Singapore in 2010, at which the Singaporeans encouraged the Netherlands' participation. The Ministry of Education in the Netherlands favoured Cito, the Dutch Institute for Educational Measurement, to take part in this project, as the issue of the assessment of 'general skills' had not been addressed from an evidence-based perspective. Since the ATC21S research project was already under way, it was decided that the Netherlands would take part only in the fourth phase of the development – the trials.

Localisation

To make the research possible in the Netherlands all material needed to be translated into Dutch. The CPS tasks were translated in October 2011, after the last corrections were made in the English versions, following the pilot results from the other participating countries. The translation was done by the experienced item developers at Cito and checked by experts at Cito for the correctness of the translations from the point of view of content. Special attention was given to the wording that is normally used in externally presented tests in the Netherlands, like final examinations and PISA, and to the language used in interactive computer applications. The experience of Cito in translating PISA tasks into Dutch was most welcome.

It was decided to concentrate in the Netherlands on the collaborative problem solving assessment tasks only. These are, generally speaking, not culturally specific and also are the most relevant with regard to the upcoming international PISA research 2015, in which problem solving will be assessed not as an individual skill but as a collaborative skill. The completion of the CPS tasks was still underway in April 2011, so this decision brought all countries onto the same timeline regarding the availability of these tasks.

The assignments that were already developed for Learning through Digital Networks were not translated for a variety of reasons. First of all, the use of an English poem as the starting point for a task using information on the internet would have been inappropriate for the Dutch situation. Developing a parallel adaptation of this task using a Dutch poem would be complicated, as relevant websites in Dutch with a comparable approach and comparable information to those used in the English version would need to be found or created. The same would apply to the task on the Arctic region. There was also no time or funding to have this transposition to the Dutch situation done. It would also have led to tasks that would differ in many aspects from the English original or the version made for Spanish, which would make an honest comparison more difficult.

As the CPS tasks involve two participants who are presented with different information on their computer screens, the correct wording of the tasks is critical in a new way. Words addressing interactive action and communication through the computer in an educational assessment context are new to test developers and students alike. A literal translation of English verbs runs the risk of producing misleading instructions because an English verb can sometimes be translated by different Dutch verbs, or expressions that have slightly different meanings. For example, 'need partner' in English can be understood as 'I need a partner' or as 'You need your partner'. Actually, the latter meaning was intended. The instructions 'retry', 'reset', and 'have another go' can all be translated as 'try again' or 'play the game again'. A correct translation is only possible when the complete instruction and the actual activities of both participants on the tasks are understood. This observation applies to the whole translation procedure. Another problem that arose was the translation of instructions or choices in which a student was required to select from three options on a bar. For grammatical reasons some Dutch wording would end in a different place from the English original, causing problems with the visual format. One practical consequence was that an allocated physical space could be too short to contain a correct Dutch translation of more than one word, since in the English original a single word could suffice. In other cases, it proved difficult to find a common Dutch equivalent for common English words. 'Transfer' can be translated literally ('Overdracht'). The English word is frequently used in Dutch, but only in the context of sports, so a more descriptive sentence had to be generated, which caused problems with layout. As mentioned, Cito's experience with PISA translations into Dutch was helpful, but as a general remark it can be observed that a literal translation can produce an artificial form of Dutch. For tasks like these, the use of common, everyday words is to be recommended, as students must, before anything else, understand what they are required to do and not be puzzled by instructions that can be understood in different ways. When translating the tasks, it is advisable to test them on small samples of different age levels and school types, asking the students if any instructions or choices are unclear or seem equivocal. The tasks themselves are already somewhat artificial in comparison to those that most students will know from computer games, so the wording (and perhaps the images as well) should aim to bridge the gap between this artificiality and the students' other experiences with instruction in games. Of course, these problems could have been addressed earlier had the Netherlands participated in the first three phases of the project.

Method

As the activities in the Netherlands were limited to the trial of the CPS tasks, the main practical issue to be addressed was to recruit schools that were willing to participate in this research project. Schools in the Netherlands are not obliged to take part in educational research projects, so the selection of schools depended on their willingness to invest students' learning time in this research. Many schools do volunteer to take part in research tests, which put the request to take part in this project in competition with other requests for research. At a national meeting of school leaders, early in 2011, those present were asked by Cito to express interest in being informed about forthcoming issues related to 21st century skills. Five primary schools and 21 secondary schools reacted to this call. This made it possible to approach these schools hoping that they would react positively to a request to take part in the trials on a voluntary basis. Four schools for primary education and eight schools for secondary were willing to have their students take part in the trials in November 2011.

Trials

The purpose of the trial was encapsulated by three questions of interest at the local level (in addition to the global project requirement for data to inform the psychometric qualities of the assessment tasks):

- Is it possible to have an online trial organised using a website situated outside the Netherlands?
- Are students across different age levels able to complete the tasks?
- How do Dutch students perform in comparison to students in other countries?

The organization of the trials was based on a design at the global level of the project in which 60 students aged 11 and 240 students aged 13 and 15 were needed. It was decided to connect age levels to grade levels in the Dutch education system: age 11 to the last year of primary education, age 13 to the second year of secondary education, and age 15 to the fourth year of secondary education (there were no restrictions on school types). In practice this probably meant that the tasks were performed by students ranging in age from 10 to 17, but with clear concentrations on the suggested age levels.

One primary school had to withdraw a few days before the trials, due to unexpected problems at their school. One secondary school withdrew at the last moment due to the Information Technology (IT) problems they were facing. According to the information given by the schools, 107 students aged 11, 200 aged 13, and 162 aged 15 would take part in the trials. Due to organizational problems, technical problems (firewall and IT problems) and other personnel issues, not all groups and not all students were finally able to take part in the trials. In the end there were 56 students aged 11, 182 students aged 13 and 119 students aged 15 who participated in the Dutch trials.

These low figures can be related to two issues of concern. First, there is the growing unwillingness of secondary schools to take part in external tests (thus exercising their right not to take part), especially when there is no direct relevant feedback to the students ('What's in it for us?'). Secondly, in some cases technical problems are encountered by schools when the need to access external websites conflicts with firewall configurations and hardware limitations at school level. In practice, it is extremely important to approach IT specialists at school level and inform them about what is going to happen. They must also have clear and detailed instructions on the timing of the trials and the system specifications which are now published on the project website (ATC21S.org). The Dutch IT specialists of the participating schools were quite willing, but even so they sometimes faced problems they did not expect, and the information given to them was often not complete.

Counterbalancing these problems, the primary schools were very eager from the start onwards. This enthusiasm related to the fact that these schools were already closely cooperating in the introduction of IT as a learning tool in their schools. They were participants in a self-initiated project to introduce Information and Communication Technologies (ICT) and 21st century skills in their didactic and instructional approaches to students. This initiative was actually directly inspired by the ATC21S project in Singapore, when representatives from the schools visited that country with a delegation of Dutch school administrators in September 2010.

Due to time restrictions, it was possible to visit only one primary school during the trial phase – a school that was very proud to be the first school in the Netherlands to take part in this research. Visiting secondary schools would have been very time-consuming and complicated, as the decision to participate in the tests was in some cases taken by those schools at the last moment. The trial took place during a convenient period for schools, from mid November to early December. This is very important, as taking part in research of this kind disrupts the regular order of things at schools, and this period is generally removed in time from those periods in which tests and school exams must be taken. The first trial was undertaken at a school for primary education. A class of 26 students, 11 years old, was split in two groups that were physically separated at two different levels in the building. After some instruction by their teachers, the students started enthusiastically with the practice task ('Light box'), which generated a buzz of discussion and excitement, but after some ten minutes the students quietened down and got involved in what they were doing.

Responses to the Trials

This first trial generated the following observations:

- Students sitting next to one another, although not working together, were sometimes seduced to exchange information, suggestions and ideas, so some uncontrolled collaboration occurred beyond the confines of the study.

- Students had a tendency to start experimenting without first exchanging suggestions and ideas. This may have been a form of contaminating behaviour as it did not appear to be related to skills in problem solving or collaboration and was apparently driven by enthusiasm and curiosity. It may have been a reaction to the circumstances of the moment or a consequence of instruction from the teacher, who spontaneously re-worded the instructions in oral form, as she would normally do. The fact that the students knew they were taking part in an experiment may have generated a buzz of excitement in which information was lost. Their behaviour showed that they had a spontaneous understanding of what to do and how to solve problems, although some lost their nerve when their partner did not (immediately) react to their communications.
- Students really liked the tasks: they were very much involved, interested in completing the tasks, and sometimes worked very quickly to do as much as possible (although that was explicitly not the goal of the tasks).
- The chat function did not allow students to decide which task to perform next if one of the pair left a task before that decision was taken.

As an experiment in its own right, it was interesting to see how schools managed to take part in the trials when they were given only electronic instructions via email. This turned out surprisingly well. On the other hand, many schools expressed disappointment about the fact that they were not given any feedback on the results.

No systematic collection of the opinions and experiences of participating students took place. This was a missed opportunity, as it would not only have given much more information on the appreciation of the tasks but would also have been a means for undertaking ongoing discussion with schools about 21st century skills, including their expectations for the future. Some coordinators from secondary schools reported spontaneously what had happened at their school. All IT problems – if present – were solved. Only in one school were the sessions postponed after the first 15 min (the reason was to be found at school level). In relation to the content, students were described as 'motivated and enthusiastic', although there was criticism of the sometimes childish or artificial level of the tasks, and the simple visual techniques involved.

Challenges

As this was the first time CPS tasks had been trialled in the Netherlands (or worldwide) it was difficult to evaluate the amount of activity and time needed to get the best information out of this trial. For the Netherlands, the investment of more time in the evaluation of the trial itself as opposed to concentration on the generation of data would have been useful. This reality demonstrates the usefulness of countries being involved in the full process rather than in the final stages as in the Netherlands' case. The schools showed a real interest in the issues of 21st century skills and in online assessment, but they wish to be more than laboratory animals. From this point of view it was a pity Dutch schools could not have been involved in the project from the very beginning.

Conclusion

The Dutch schools that took part in this research project all showed that schools are ready to embrace tasks like CPS, both from the point of view of content and from the point of view of IT. Regretfully, so far, this research project has been an incident in Dutch education, with no practical results for the schools that took part and no politically supported follow-up yet. By making the issue of 21st century skills more visible through the issue of assessment, the project may have encouraged the schools that took part in the research to invest time to teach the skills in a more structured way. However, at a governmental level more attention is currently given to raise the levels of basics skills such as language and arithmetic. Since the Dutch government introduced the precursors of 21st century skills – then named 'general skills' – into education some 20 years ago, it may be time to pay more structural attention to the teaching and learning and to the assessment of these skills. This need, however, must be seen in the context of the difficulty of assessment of these skills and of current political perspectives on accountability. The issue of 'objective' accountability and of international comparability is limiting the possibilities for the innovation needed in education. Although the Netherlands have been in the forefront of adjusting the education system to the challenges of contemporary society, the country might now suffer from the so-called 'Law of the handicap of a head start' (in Dutch: 'Wet van de remmende voorsprong'), coined by the Dutch historian Jan Romein in 1937. This 'law' indicates that getting a head start in a given area may become a handicap in the long-term, as early success might blind you for opportunities later on when others enter the same arena. The introduction of 'general skills' in Dutch education some 20 years ago was ahead of the international discussions on 21st century skills. But the results of the introduction and implementation of the 'general skills' have not been evaluated profoundly, nor is there a feeling that these skills need a boost. From the perspective of the ATC21S project, the role of information technology in particular, in both learning and assessment, should urgently be addressed in Dutch educational policy, as the developments in IT have been enormous since the time the 'general skills' were introduced.

Part V
Implementation at Classroom and System Levels

This part describes implementation of the ATC21S system at the classroom level. Woods, Mountain and Griffin (2015, Chap. 14) explore the professional development modules, the reports provided for teachers and how teachers interpret the reports to identify intervention strategies at the classroom level. Adamson and Darling-Hammond (2015, Chap. 15) describe the consequences and implications for systems of education of the ATC21S approach to assessment, scoring, reporting, teaching and professional development. Their discussion focuses on the implications of taking the process to scale for education systems. Chap. 15 follows from Darling Hammond's (2012) discussion in Volume 1 of the series in which policy implications for 21st century education systems were explored. This chapter explores issues associated with infrastructure, technology, policy, and the willingness of systems to implement the assessment of 21st century skills.

Woods, K., Mountain, R., & Griffin, P. (2015). Linking developmental progressions to teaching. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills: Methods and approach* (pp. 267–292). Dordrecht: Springer.

Adamson, F., & Darling-Hammond, L. (2015). Policy pathways for twenty-first century skills. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills: Methods and approach* (pp. 293–310). Dordrecht: Springer.

Chapter 14
Linking Developmental Progressions to Teaching

Kerry Woods, Roz Mountain, and Patrick Griffin

Abstract This chapter presents an approach to teaching 21st century skills on the basis of results from the assessment tasks described in Chaps. 3 and 4. Teachers' understanding of the skills being assessed and described in the developmental progressions is essential if they are to implement learning and teaching activities in their classrooms across the curriculum. In this chapter the interpretation of three report formats for the formative assessment of 21st century skills is explained in the context of a developmental learning framework. A teaching approach is outlined and examples of teaching and learning sequences are presented. The activities and tasks in the examples are adapted for students at different stages on the progressions, and suggestions are given for targeting teaching and learning to the stages of students on a combination of progressions.

A Developmental Approach to Assessment and Learning

In this section, the basis of a developmental learning approach to using assessment for teaching 21st century skills is outlined. The aim of the approach is to move a student's learning forward along a path or progression of increasingly complex knowledge and capabilities. The focus is on recognition of a student's readiness to learn and the process of building upon the current stage of learning. By contrast, a deficit approach to assessment and teaching focuses on discovering the things that a student cannot do, and teaching is then designed to address those deficits. A developmental approach to learning assumes that there is a typical pathway that describes and maps the progress of a student through stages of increasing knowledge, skills and understanding. In the context of the ATC21S™ Project[1] assessment materials, developmental assessment and learning is based on

[1] The acronym ATC21S™ has been globally trademarked. For purposes of simplicity the acronym is presented throughout the chapter as ATC21S.

K. Woods • R. Mountain • P. Griffin (✉)
Assessment Research Centre, Melbourne Graduate School of Education, University of Melbourne, Parkville, VIC, Australia
e-mail: p.griffin@unimelb.edu.au

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_14

aspects of the theory and application of three academics: Lev Vygotsky, Robert Glaser, and Georg Rasch (Griffin 2007).

Lev Vygotsky (1978) proposed the concept of the zone of proximal development (ZPD), which can be thought of as an ideal space in which people learn most effectively. People learn effectively within their ZPD because they have enough prior knowledge to scaffold their learning of more complex skills or information, but not so much as to lead to disengagement because they are being taught concepts or material that are too simple. Vygotsky is also well known for his theory that social interaction plays a fundamental role in children's cognitive development.

The ZPD is defined as the zone between a student's actual developmental stage and the level that is beyond their current capacity when working independently, but not when supported by a more able or knowledgeable person (Vygotsky 1978). Between these two levels is the zone in which the student can succeed on a learning task or activity with the help of an adult or more capable peer. Vygotsky explained that students vary in their actual developmental stage, involving skills they demonstrate independently, as well as in their zone of proximal development – the distance between the actual developmental stage and the potential developmental stage with adult or mentor guidance. In his words:

> The zone of proximal development defines those functions that have not yet matured but are in the process of maturation … These functions could be termed the 'buds' or 'flowers' of development rather than the 'fruits' of development. The actual developmental level characterises mental development retrospectively, while the zone of proximal development characterises mental development prospectively.
>
> Vygotsky 1978, p. 86

One way to apply Vygotsky's insight is to acknowledge that teaching and learning should be informed by students' emerging skills, and seek to strengthen or extend these, rather than focusing on skills and abilities that are already established.

Robert Glaser was a researcher who studied aptitude, testing in education, the use of technology in education, and tailoring instruction to individuals. He introduced the term 'criterion-referenced interpretation' to help us understand assessment data in terms of the skills that the students demonstrate (Glaser 1963). Criterion-referenced interpretation of assessment data describes the performance of an individual as a skill, or set of skills, rather than simply as a number, percentage, or comparison with other students who have completed the same assessment. Student proficiency is mapped to a skill or behaviour criterion (or set of criteria) to give meaning to the set of capabilities a student can demonstrate (Griffin 2007).

Georg Rasch, a Danish mathematician, made an important contribution to psychometrics that has been applied to the measurement of knowledge, abilities, and attitudes. Using latent trait theory (i.e., constructing measures of variables that are not directly observable) and mathematical modelling, Rasch (1960/1980) was able to formally measure the location of student ability and test item difficulty together on a single scale. In the simple Rasch model, the probability of a correct response on a yes/no test question is a function of the student's position on the scale relative to the difficulty of the item. Both the difficulty of test items and the ability of students can be estimated from assessment data using specialised computer programs.

Scale	Students	Test items	Progression levels
6	High x	Difficult	
	ability	items	
5	x		Level 6
	x		
4	x		
	x	11 14 17	
3	xxx	15	
	xxx		
2	xxxx	5 8	Level 5
	xxxxx	2 23 10	
1	xxxx	1 24	
	xxxxxx	13 26	
0	xxxxxxxx	9 12	Level 4
	xxxxxxxx	21	
-1	xxxxxxx	3	
	xxxxxx	7 15 16	Level 3
-2	xxxxxx	19	
	xxxx	2 4 27 28 30	Level 2
-3	xxxx		
	xxx	6 22	
-4	xx		
	xx	20	
-5	x		Level 1
	Low	Easy items	
	ability		
-6	x	18	

Fig. 14.1 Map of the distribution of task difficulty, student abilities, and derived levels of proficiency interpreted from clusters of items of similar difficulty

Where a student's ability and the difficult of a test item are aligned in the estimation, the probability of the student answering that item correctly is 50 %. This modelling can then be used to support the interpretation and empirical validation of levels of increasing competence along a developmental progression (Griffin 2007).

To explain this in a little more detail, the output of a Rasch modelling analysis can be shown as a variable map, an example of which is presented in Fig. 14.1.

The *X*s on the left of the scale represent the position of students on an ability continuum ranging from low to high ability. The numbers on the right of the scale indicate the position of test items along a continuum from low to high difficulty.

Combining the insights of Vygotsky (1978), Glaser (1963) and Rasch (1960/1980), a student's assessment score can be interpreted in terms of performance criteria that are grouped into a stage of competence (Griffin 2007). This information can be used to determine a point of intervention where learning can be scaffolded for a student or group of students. The challenge for educators is to identify students' emerging skills, the skills located within their ZPD, and provide the right level of support at the right time (Griffin). The implications for teaching and learning practice are that test scores are no longer simply an end-point – a piece of summative information from the past that describes the skills or information students have retained – nor a means of comparing students with each other. Instead, test scores can be interpreted in terms of the skills a student is beginning to develop to provide the starting point for planning instruction (Griffin). So, as can be seen in the example shown in Fig. 14.1, sets of skills with similar levels of difficulty can be grouped into levels along a progression and interpreted in terms of their commonalities. In other words, we can ask ourselves 'What do the skills that cluster at this broad level of difficulty have in common?' Or 'How do the skills at this level along the developmental progression differ from those that are more (or less) difficult?' These questions can help us to understand and interpret what we mean when we talk about the development or unfolding of skills and understanding in a particular area of learning.

Using Developmental Frameworks to Describe and Understand Learning

Many developmental frameworks are general in nature and can be applied to a range of situations. One example familiar to many teachers is Bloom's Taxonomy (Anderson et al. 2001), which provides a classification framework with six stages of increasing competence as shown in Table 14.1.

As an example of the application of Bloom's taxonomy to a practical skill, students who have learned about tagging in social media, twitter or other online contexts may demonstrate increasing stages of competence as shown in Table 14.2.

Students at the lowest stage who create and use tags to follow comments on a topic and post their own comments (Stage 3) are ready to learn to organise tags effectively (Stage 4) and need to be given tasks to help them understand and practise organising tags for a variety of purposes. Students who are able to organise tags

Table 14.1 Summary of the cognitive dimension of Anderson and Krathwohl's revision of Bloom's taxonomy (Anderson et al. 2001)

1 Remembering	Retrieving relevant knowledge from long-term memory.
2 Understanding	Determining the meaning of instructional messages, including oral, written and graphic communication.
3 Applying	Carrying out or using a procedure in a given situation.
4 Analysing	Breaking material into its constituent parts and detecting how the parts relate to one another and to an overall structure or purpose.
5 Evaluating	Making judgments based on criteria and standards.
6 Creating	Putting elements together to form a novel, coherent whole or make an original product.

Table 14.2 Example of an application of Bloom's taxonomy

Stage	Example behaviour to demonstrate knowledge of tagging in social media or other online contexts
3 Applying	Creates and uses tags to follow or post comments on a particular topic
4 Analysing	Able to organise tags for a given purpose
5 Evaluating	Comments on the usefulness of tags and optimises effectiveness for a purpose

effectively (Stage 4) are ready to learn to evaluate tags in terms of their purpose and suggest improvements and adjustments (Stage 5). This may involve exploring different perspectives, and looking at ways to organise information and communicate it with awareness of a range of possible perspectives.

In contrast to classification taxonomies, which are commonly drawn from theory, empirical progressions are derived through an analysis of assessment data from large numbers of students. Statistical methods are used to determine the order of acquisition of skills, knowledge or attitudes based on a sample of students, as illustrated in Fig. 14.1 above. Empirical progressions represent a typical pathway for students' development and can therefore be a useful reference for teachers to set goals for their students and to plan tasks designed to engage and challenge them.

Using empirical data and theoretical frameworks, the ATC21S project produced progressions of learning for students. Summary versions of the progressions that describe students' collaborative problem solving skills (Hesse et al. 2015) are presented in Tables 14.3 and 14.4. Drawing from the work of Wilson and Scalise (2015), Tables 14.5, 14.6, 14.7, and 14.8 describe hypothesised progressions of skills linked to learning through digital networks and Table 14.9 shows an overarching progression for these as a skill associated with Information and Communication Technology skill development. This is referred to in this volume as LDN-ICT. In each progression, stages of competence can be identified by a sequence of letters or numbers, or short descriptions that summarise the overall theme for each stage. Examples of each of the progressions and labels are shown below.

Using a Developmental Model to Plan Teaching

Assessment data interpreted within a developmental framework can be used to understand how students progress from one level of competence to the next (Griffin 2007). In this section we will be introducing the assessment reports that can be derived from the ATC21S tasks. These reports can be used by teachers to plan for instruction and to organise their classes, and also by students to review their progress on each of the progressions shown above.

Table 14.3 A progression of cognitive skills for collaborative problem solving

Skill level	Cognitive skills for collaborative problem solving
F	**Refined strategic application**: Students' sequential investigations and systematic behaviour require fewer attempts for success and are completed in an optimal amount of time. The student works with their partner to identify and use only relevant and useful resources. The student has a good understanding of the problem and can reconstruct and/or reorganise the problem in an attempt to find alternative solution paths.
E	**Efficient working**: Students' actions appear to be well thought out, planned and purposeful, identifying the necessary sequence of sub-tasks. They identify cause and effect, base their goals on prior knowledge and use suitable strategies to gain a correct path solution for both simple and complex tasks. The students can modify and adapt their original hypotheses, in light of new information, testing alternative hypotheses and adapting additional or alternative ways of thinking.
D	**Strategic planning and executing**: Students can identify connections and patterns between multiple pieces of information. They are able to simplify the problem, narrow their goal focus and increase co-working by planning strategies with their partner. The students adopt strategic sequential trials and increasingly display systematic exploration. They can successfully complete sub-tasks and simpler tasks.
C	**Sharing and connecting information**: Students recognise the need for more information, realising that they may not have all the required resources, and allocate their own resources to their partner. They attempt to gather as much information as possible and begin connecting pieces of information.
B	**Establishing information**: Students identify possible cause and effect of actions, demonstrate an initial understanding of the task concept and begin testing hypotheses and rules. They limit analysis of the problem, using only resources and information to hand. Student goal setting is limited to generation of broad goals.
A	**Exploration**: Students explore the problem space but exploration is limited to following instructions, adopting a singular approach, and focusing on isolated pieces of information. Trial and error appears random and there is little evidence of understanding the consequences of actions, resulting in a lack of progress through the task.

Report Formats

Reports can be generated on completion of the ATC21S assessment tasks, to place students at levels or stages on progressions of skill and understanding. One of these reports is a *learning readiness report*. An example is provided in Fig. 14.2. This report shows a series of learning stage descriptions in a particular domain – e.g. cognitive skill for collaborative problem solving (Hesse et al. 2015) – arranged from the lowest stage at the bottom to the highest stage at the top. This can be linked to the progression of skills described in Table 14.3.

The learning readiness report summarises the capabilities that a particular student is currently developing in a given domain, and those that the student might be expected to develop next, and thus can be used to identify an appropriate focus for student learning and teaching intervention. It can be used by students as feedback

Table 14.4 A progression of social skills for collaborative problem solving

Skill level	Social skills for collaborative problem solving
F	**Cooperation and shared goals**: Students work collaboratively through the problem solving process and assume group responsibility for the success of the task. Feedback from partners is incorporated and used to identify solution paths or modify incorrect paths. The students can evaluate their own and their partners' performance and understanding of the task. The students may tailor their communication and manage conflicts with partners successfully, resolving differences before proceeding on a possible solution path.
E	**Appreciated and valued partnership**: Students actively participate in scaffolded and unscaffolded environments. The students initiate and promote interaction with their partners and acknowledge and respond to contributions from their partners. Despite efforts, differences in understanding may not be fully resolved. The students are able to comment on their partners' performance during the task.
D	**Mutual commitment**: Students persevere to solve the task as shown by repeated attempts and/or use of multiple strategies. They share resources and information with their partners and modify communication where necessary to improve mutual and common understanding. Students have an awareness of their partner's performance on the task and can comment on their own performance.
C	**Awareness of partnership**: Students become aware of their partner's role in the collaborative problem solving process and recognise the need to engage with their partner. They discuss the task with their partner and make contributions to their partner's understanding. The students report to their partner regarding their own activities on the task.
B	**Supported working**: Students actively participate in the task when it is scaffolded but work largely independently. Communication between partners occurs but is limited to significant events and information necessary to commence the task.
A	**Limited interaction**: Students commence the task independently with limited interaction from partner, mainly prompted by instructions. They may acknowledge communication cues by their partner but have not started to work collaboratively. Most communication occurs at the beginning of tasks and only in those tasks where the instructions are clear.

Table 14.5 An hypothesised progression of LDN-ICT skills as a consumer in networks

Skill level	LDN-ICT literacy: consumer in networks
High	**Discriminating consumer**: Students are able to seek expert knowledge through networks and judge the credibility of sources/people. They filter, organise, manage, evaluate and reorganise information into an integrated and coherent knowledge framework. They select optimal tools for tasks and tailor searches and interactions to their own and their audience's circumstances.
Medium	**Conscious consumer**: Students construct targeted searches, select appropriate tools and strategies, compile information systematically and are aware that credibility of sources is an issue.
Low	**Emerging consumer**: Students perform basic tasks in a network environment, searching for information using common search engines. They have some knowledge of social media tools.

Table 14.6 An hypothesised progression of LDN-ICT skills as a producer in networks

Skill level	LDN-ICT: producer in networks
High	**Creative producer**: Students produce attractive digital products, selecting from multiple technological options and tools to best suit the purpose. They are able to assemble digital products creatively through a process of assembling distributed contributions. Students make use of their understanding of skills in a team to make best use of available expertise.
Medium	**Functional producer**: Students establish networks and communities and organise communication within these networks using appropriate tools and styles. They plan and develop creative and expressive websites, blogs or games, with an awareness of security and ethical and legal issues. Their work is based on established models.
Low	**Emerging producer**: Students produce simple representations of information from templates. They are able to use a computer interface to post an artefact and start an identity.

Table 14.7 An hypothesised progression of LDN-ICT skills in building social capital in an online environment

Skill level	LDN-ICT: developer of social capital
Very high	**Visionary connector**: Students take a cohesive leadership role in building a social enterprise. They reflect on experience in social capital development.
High	**Proficient connector**: Students initiate opportunities for developing social capital through networks. They encourage multiple perspectives and support diversity in networks.
Medium	**Functional connector**: Students are aware of multiple perspectives in online social networks. They contribute to building social capital through a network, and encourage participation and commitment from others.
Low	**Emerging connector**: Students are aware of online social networks, and participate as observers or passive members, or engage actively at a basic level in social enterprises.

Table 14.8 An hypothesised progression of LDN-ICT skills in building intellectual capital in an online environment

Skill level	LDN-ICT: developer of intellectual capital
Very high	**Visionary builder**: Students question existing social media architectures and develop new architectures. They engage in dialogue at the interfaces between social and knowledge building architectures.
High	**Proficient builder**: Students understand and make use of various architectures in social media (tagging, polling, modelling, role playing) to link to knowledge and expertise. They choose optimal tools to locate and access information. They interrogate data for meaning and distinguish between relevant and extraneous information. Students create, share and reframe mental models to build collective knowledge.
Medium	**Functional builder**: Students are aware of multiple perspectives in knowledge organisation. They are able to organise tags thoughtfully. Students understand the mechanics of collecting and assembling data to create a shared representation. They know when to draw on collective intelligence.
Low	**Emerging builder**: Students are able to make tags or post a question online, and have some knowledge of survey tools.

Table 14.9 An hypothesised overarching progression of LDN-ICT skills

Skill level	Overarching LDN-ICT developmental progression
E	Students can successfully navigate the web and efficiently select relevant resources and materials and apply these appropriately to tasks. The students can reflect on their overall performance on tasks. They take an active role in leading their team to successful completion of tasks.
D	Students can distinguish and sort between relevant and irrelevant statements relating to content. They can provide explanations for a change in answer based on partner feedback. They are also able to reflect on their own and their partner's performance. They can create their own materials and incorporate them into existing interfaces.
C	Students generate new ideas relating to content using available tools. They are able to upload appropriate images, audio and word documents correctly. Students can produce an accurate pie chart by analysing data online and in a graph. They can generalise from website content to generate hypotheses and questions relating to content. They can suggest appropriate website addresses with relevant, preferred content.
B	Students can sort information by relevance and select the relevant web link for the current task. They can forage/gather and analyse appropriate information from websites to enable them to answer questions. Students can answer questions on the content of tasks and provide an explanation for a previous answer or action. They can create simple representations of their ideas using available tools.
A	Students are able to use available simple tools, such as drawing tools and icons, to drop and drag and create pictures/landscapes They can copy text from one location and paste into another. They can amend existing content on a page and generate new basic content using available tools. Students can access the resources available to them to search for pieces of information including available web links, although not always those relevant to the current task. They can engage with available help-podcasts when requiring further instructions. The students can follow simple instructions as well as activate content on a page.

on their current skills and understanding in that domain, and to provide information on how to improve their knowledge and performance. It can be used by teachers to support their plans for future learning experiences for students.

The black bar on the report shows the student's stage of learning readiness on the progression. The associated description of the stage outlines the skills that the student is currently ready to learn. The student's estimated stage on the report is not an achievement level but, rather, a point of intervention that teachers can use to make decisions about the best possible learning program for the student and to set goals and intentions for teaching and learning.

The position of the black bar within a stage indicates whether the student is just beginning to develop the skills, consolidating the skills, or moving towards mastery and ready to start a new stage of learning. As students move to the upper half of a stage, it can be helpful to look ahead to the next stage on the progression and to reflect on the sorts of skills and capabilities the student is working towards. With the additional support of scaffolding or modelling, teachers and students can use this information as a way of setting more challenging targets for learning.

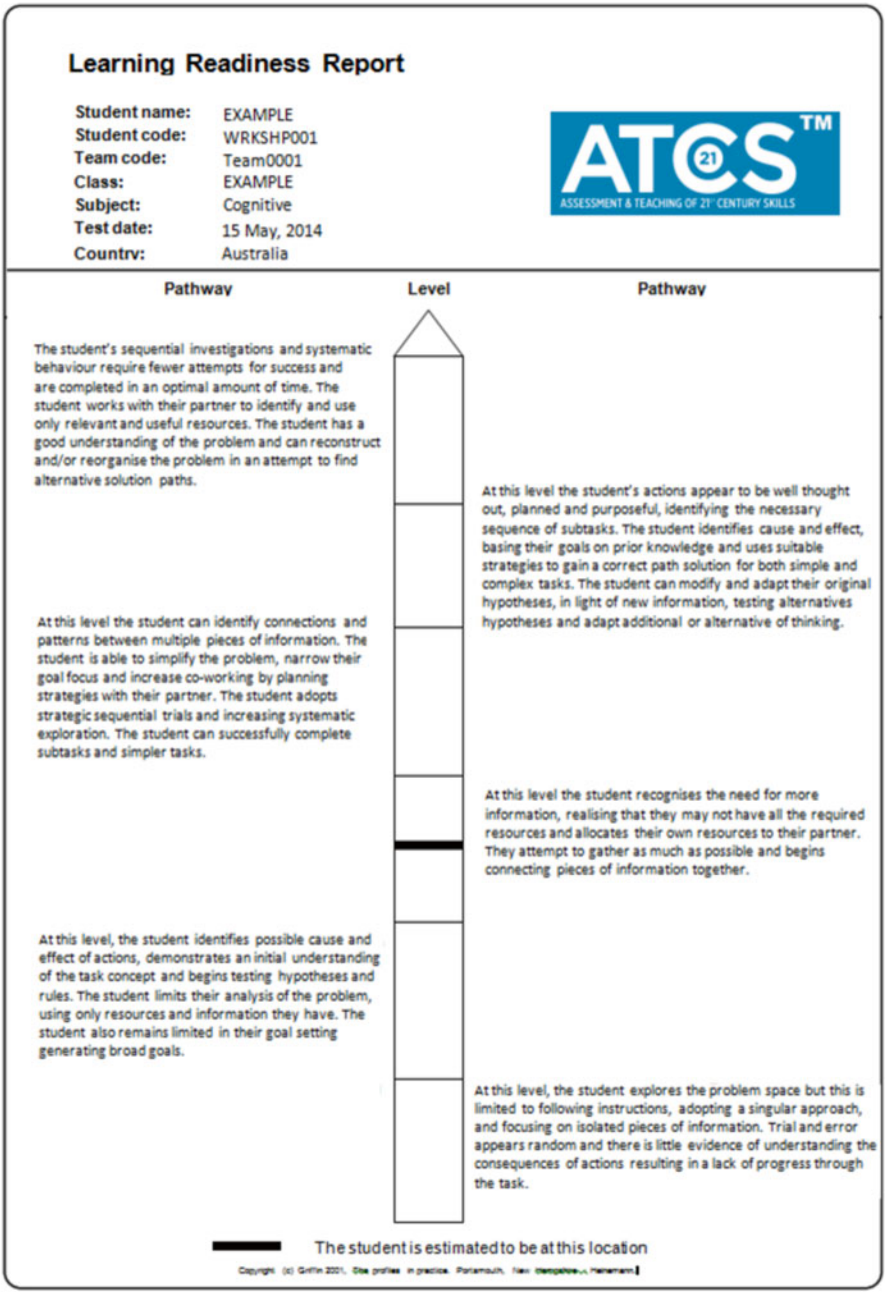


Fig. 14.2 Example of a learning readiness report

Another report that can be generated from the ATC21S tasks is the *student profile report*, which maps an individual student's stage of learning across a number of learning domains (i.e., cognitive and social skills in collaborative problem-solving, and ICT literacy skills). This report, an example of which is shown in Fig. 14.3, is designed to support consideration of a student's individual pattern of strengths and abilities.

There is no expectation that a student will be at the same stage across all learning domains simultaneously; nor is there an expectation that he or she will move through different stages at the same general rate of progress. Indeed, it is quite common for a student to be working at a high overall stage in the social aspects of collaborative problem solving while working at a lower stage in the cognitive aspects. In contrast, some students may have particular strengths in the cognitive aspects of problem solving but struggle to develop the social aspects needed to be a skilful collaborative problem solver.

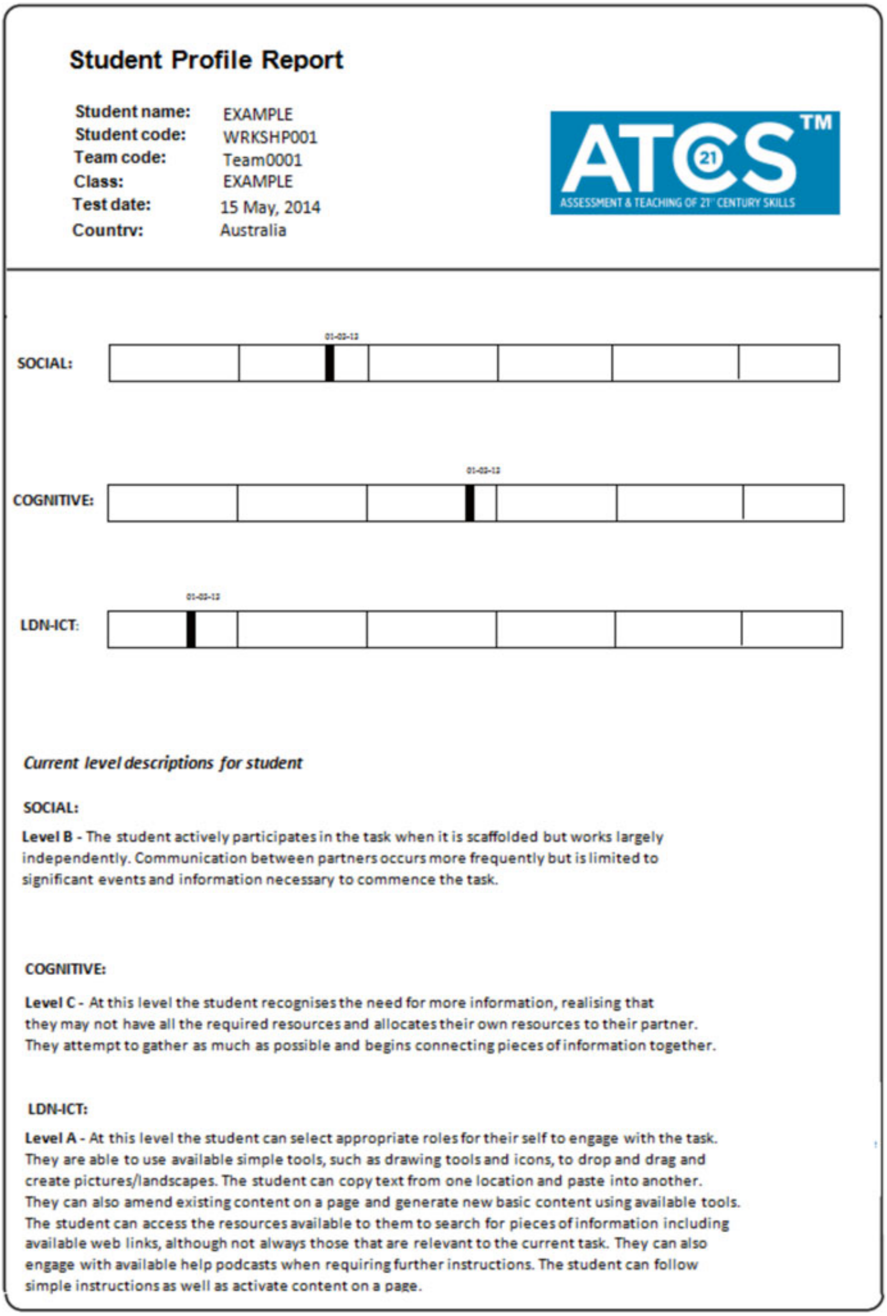
Reviewing Student Progress

Students and teachers can work together or independently to review progress and set targets for future learning. The learning readiness report can be used to confirm understanding of what a student can do with confidence and what they are ready to start learning with scaffolding, modelling, or the support of a more capable other. The profile report can promote understanding of a student's particular pattern of strengths and abilities.

Teachers can use this information combined with other evidence from work samples, classroom observation and other assessments to develop a rich understanding of students' knowledge and skills. This allows teachers to formulate a set of learning intentions that will engage and challenge each student. Through an understanding of students' learning preferences and interests, teaching can be adapted to promote student progress.

In determining learning intentions for students working at the same generalised stage of skill and understanding, it is important to consider both long term and short term goals and plans. Learning intentions should be clear and achievable so students understand what they need to do or demonstrate in order to make progress. Developmental learning progressions are very useful in this process, as they describe the skills and abilities at both the students' current stage and the next stage. Once student learning intentions have been agreed, the next step is to use this information to plan appropriate learning activities. Over time, teachers can build up a bank of successful strategies and learning experiences for students at each developmental stage. Some ideas and examples suggested by teachers are provided in the next section.

As part of the process of planning and reflection, teachers and students may make notes about the evidence of learning that they expect to be able to observe. This will allow for effective review of the goals that are set as well as the teaching



Student Profile Report

Student name: EXAMPLE
Student code: WRKSHP001
Team code: Team0001
Class: EXAMPLE
Test date: 15 May, 2014
Countrv: Australia

Current level descriptions for student

SOCIAL:

Level B - The student actively participates in the task when it is scaffolded but works largely independently. Communication between partners occurs more frequently but is limited to significant events and information necessary to commence the task.

COGNITIVE:

Level C - At this level the student recognises the need for more information, realising that they may not have all the required resources and allocates their own resources to their partner. They attempt to gather as much as possible and begins connecting pieces of information together.

LDN-ICT:

Level A - At this level the student can select appropriate roles for their self to engage with the task. They are able to use available simple tools, such as drawing tools and icons, to drop and drag and create pictures/landscapes. The student can copy text from one location and paste into another. They can also amend existing content on a page and generate new basic content using available tools. The student can access the resources available to them to search for pieces of information including available web links, although not always those that are relevant to the current task. They can also engage with available help podcasts when requiring further instructions. The student can follow simple instructions as well as activate content on a page.

Fig. 14.3 Example of a student profile report strategies. By specifying what to expect in students' classroom behaviour, teachers are able to identify the point at which students have moved to a new stage of understanding and are ready to take on new challenges in their learning.

Generalising Intervention and Differentiating Instruction: Class Reports

Teachers can also refer to a report that plots the current learning readiness of all students in a class and use this to make decisions about ways to organise small-group learning experiences or to foster mentoring by pairing less able students with more able students. Often the same or similar learning intentions can be used for groups or clusters of students who are working at the same generalised stage of proficiency. In many classrooms, teachers can expect to have students working at two, three, four, or even five stages of learning readiness. One way to visually determine groups of students operating at the same stage is to use a *class report*. An example is provided in Fig. 14.4.

The class report can thus be used to help teachers differentiate instruction to best meet the learning needs of students who are working at different stages of understanding or proficiency. Some suggestions for the practice of differentiated instruction are provided in detail below.

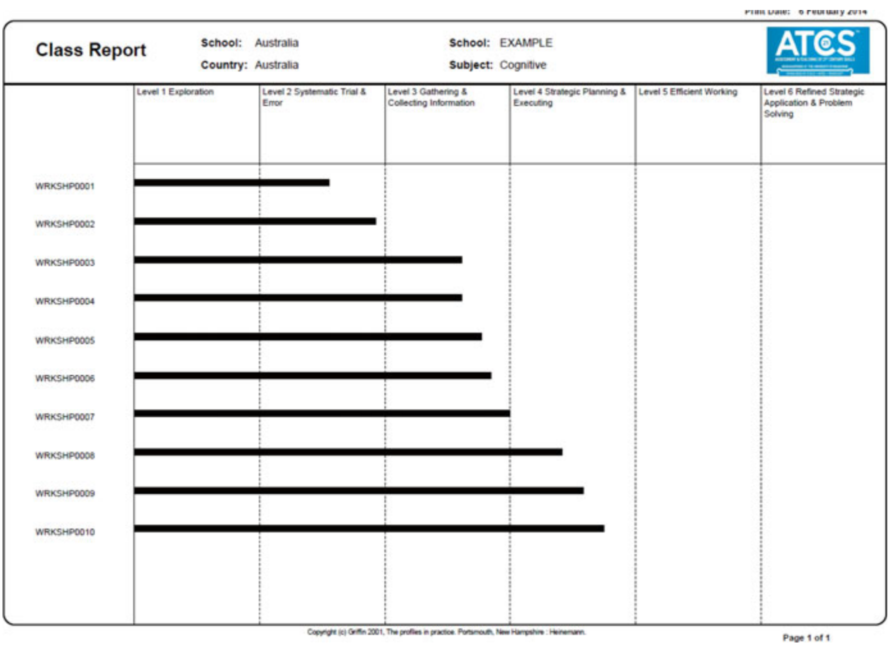


Fig. 14.4 Class report showing students at different stages of skill and understanding

Teaching 21st Century Skills

By definition, 21st century learning tasks can be open-ended, involve unbounded sets of information, and may involve on-going redefinitions of the goal of the task. It is important that students develop skills to establish and adapt goals according to available information, seek out relevant and valid information for the task, and continually monitor their own progress. The teacher's role is to set highly motivating tasks with achievable goals and to provide sufficient structure and scaffolding based on a thorough understanding of the students' interests and needs. The students also set goals and targets for their own learning, and move forward with a clear understanding of the usefulness and application of the new skills and understanding they are developing.

Another approach that fits well with the teaching of 21st century skills is tailored and differentiated instruction. Tomlinson and McTighe (2006) noted that teachers in differentiated classrooms draw upon strategies such as small-group instruction, materials presented at a variety of reading levels, personalised rubrics, learning contracts, a variety of product and task options with common learning goals, and independent studies. Small-group instruction may be particularly helpful in targeting learning tasks, allowing students to shape their own learning goals and to seek out and select materials and information of relevance to the task. The task of teachers is to provide the most effective structure through the establishment of smaller groups based on similar abilities or to provide opportunities for peer mentoring.

Teaching and Learning in Mixed Ability Classrooms

This section provides an example of the way teachers can use student assessment data as a foundation for planning a targeted and differentiated teaching sequence. The assumption is that the example teaching sequence provided here is suitable for delivery across multiple lessons and within a mixed ability class of students. In other words, it is expected that students in the same classroom may be working at any of four, or possibly more, stages of knowledge and understanding in the cognitive and social aspects of collaborative problem solving or the development of social and intellectual capital. Table 14.10 represents this for a hypothetical classroom in which students are spread across four stages of proficiency in each of the cognitive and social aspects of collaborative problem solving.

A teacher may target learning experiences similarly for a group of students to improve their performance on the cognitive aspects of collaborative problem solving, but one or two in the group may benefit from different support or conditions to build their capacity to work with others on the task. The example of a teaching and learning sequence presented below suggests ways that a teacher in a class might differentiate instruction for students working at different stages of cognitive skill while also being mindful of the students' different stages of social skill and understanding. In other words, teachers may need to think about their students' competence across two skill areas simultaneously.

Table 14.10 Illustration of the spread of students across cognitive and social aspects of CPS in a hypothetical class of 20 students

Learning stage	Cognitive stage A	Cognitive stage B	Cognitive stage C	Cognitive stage D
Social stage A	XX	X	–	–
Social stage B	XX	XX	X	X
Social stage C	X	XXX	XX	X
Social stage D	X	X	X	X

Note: Each student in this illustrative table is represented by an X

Application to Teaching and Learning Practice

Taking the empirical information that has been generated through the ATC21S project on the skills identified as crucial for students to learn, it is important to demonstrate in a practical way the feasibility of using this information for the purpose of teaching and learning. If the skills can be assessed but are not teachable, the value of the assessments is questionable to say the least. The remainder of this chapter describes a preliminary process to explore the feasibility of using ATC21S assessments to promote growth in students' skills along empirically derived developmental progressions such as those set out in the progression tables (i.e., Tables 14.3, 14.4, 14.5, 14.6, 14.7, 14.8, and 14.9 above).

This section describes ideas and examples for teaching ATC21S skills that were generated by teachers, ICT coordinators and curriculum specialists working with some of the ATC21S developmental progressions as their guides. The brief provided to these teachers, coordinators and specialists was to design teaching and learning sequences that could be implemented by teachers to best promote learning for students at each stage on the progressions. Two examples are described below, showing how a common theme can be used to design a task that can be differentiated for each stage on the progression.

Example: Collaborative Problem Solving – Cognitive Skills

Teaching and Learning Intention

In this section, a teaching and learning sequence is presented with the purpose of building students' capacity to take a structured approach to planning. This includes:

- identifying information they require but have not been given;
- searching for and collecting information they need;
- organising information they are given together with information they collect;
- following a process to generate ideas, present and discuss them, and finally to decide on a single idea to follow through;
- testing ideas for feasibility relating to a given set of constraints; and
- presenting a plan in sufficient detail to be implemented by another group.

Theme: Planning an excursion. Students are given the task of working in groups to discuss and agree on a plan for an excursion with given time and budget constraints. Student groups present their plans to the class, after which the class can vote and, if practical, decide to go on the most popular excursion. The class should evaluate the plans against other criteria, such as feasibility, clarity and creativity. If the task is not suitable for the school or class context, other ideas for themes can be used and developed in a similar way. Some examples are: planning a school event or a celebration and preparing a presentation on a particular topic.

In the case of an excursion, the plan should include:

- mode of transport;
- itinerary, including schedule of times;
- schedule of costs;
- parent permission forms;
- email to school principal requesting permission; and
- booking requests (e.g. email) for any attractions/museums, and so on, to be visited.

The goal of the activity can be adapted to address a particular topic to be covered in a curriculum or subject area. Some examples are given in Table 14.11.

Differentiation for Each Stage on the Cognitive Skill Progression

To differentiate teaching for students at different stages of skill or understanding, teachers may refer to their general location on a progression such as the one illustrated in Table 14.3 (Hesse et al. 2015). To briefly recap, that progression describes

Table 14.11 Examples of tasks embedded in curriculum subject areas

Subject area	Goal of the excursion
Science Grade 4	Investigation of an ecosystem of a given type, or containing given elements. Students could be expected to create a model of the ecosystem showing all the interactions, and to give examples of evidence of the interaction.
Art Grade 7	Exploration of different styles of painting. Students could be directed to find two typical examples of paintings in each of a given set of styles and explain why they represent the given style. Students could write a description of each painting and include background information on the artist. Some examples of styles that could be explored are:
	Cubist
	Impressionist
	Modernist
	Expressionist
History Grade 8/9	Compilation of biographical information on a particular historical figure with local relevance to the city or country. The biography could be presented in the form of a play, documentary or book.

stages of increasing competence as students move from an exploratory or trial and error approach (Stage A) to an approach based on establishing information (Stage B) to approaches that demonstrate capacity to share and connect information (Stage C) to strategically planned and executed approaches (Stage D) then those based on efficient working (Stage E) and refined strategic application (Stage F). This can be used to develop expectations about the next step for students who are working at different points along a learning continuum.

In the practical example described above, students who are just beginning to develop some initial planning skills will be learning to discuss ways of planning within a group to direct their activity in completing the task. They will be able to think of ideas and test them out through discussion or experimentation, but they require some guidance on taking a more directed approach to understanding the resources they are given or are able to gather.

As students improve their understanding and proficiency, they will begin to select their own methods and tools to structure their planning process. They need to be supported in developing an ability to organise information from different sources, which may not coincide, and to reorganise information in the light of additional or altered information.

For students who have well-developed planning skills, the activity should promote a deeper understanding of generating creative ideas by combining contributions from all members of the group, and understanding the different strengths and abilities within a group. Students should be supported to develop increasing levels of sophistication in targeting communication online and face-to-face in order to:

- search for information;
- encourage contributions from all group members; and
- challenge group members.

The section below presents detailed ideas and targeted strategies for working with students at different stages of proficiency.

For students working at Stage A on the cognitive skills progression, who typically use an exploratory, trial and error approach:

- Present this activity as a closed task by naming a destination with two feasible modes of transport by which it can be reached. Provide instructions in verbal, written and/or pictorial form. Scaffold students' searching and planning by providing two or three websites, a map and a few relevant and irrelevant public transport timetables. Work collaboratively with the group to investigate the resources at their disposal and, if required, to establish some initial rules for deciding which resources are likely to be useful.
- Leave time at the end of the session for students to describe the process they went through, to name one thing they contributed, and one thing each other member of the group contributed.
- Provide a rubric for students to evaluate, against specific criteria, the excursions proposed by the class. This could be done after the students have an opportunity to vote for the idea they like the most, to contrast these two methods of evaluation.

For students working at Stage B on the cognitive skills progression, who are typically identifying possible cause and effect relationships and beginning to test hypotheses and rules:

- Build in an explicit planning stage for students to identify and list the information they need.
- Scaffold the search for modes of transport by explicitly teaching internet searching skills, including ways to restrict searches and evaluate the usefulness of different sources of information.
- Provide a structure for negotiations to select a single idea for an excursion. This could take the form of a set of guiding questions. Each student could present one or two ideas to be typed up in a single document. Direct the students to use a colour coding system to indicate the stage of negotiation of the ideas, for example:
    - Green for proposed ideas
    - Yellow for ideas that have been discussed
    - Blue for ideas that have been agreed upon for final voting
- Allow time for reflection to identify the process they followed, what worked well, and what they could have done differently.
- Ask students to reflect on and describe how the skills they learned in doing the task could be applied to other areas of study or life.
- Conduct a class discussion on how the excursion plans could be evaluated and facilitate agreement on the criteria to be used. Following an evaluation against the agreed criteria, allow students to vote for the excursion they would like to go on.

For students working at Stage C on the cognitive skills progression, who are typically learning to share and connect information:

- Draw up a list of requirements for the excursion, two of which are challenging to satisfy in one excursion (i.e. requirements that are contradictory to each other).
- Specify a planning stage for students to decide on a method for generating and selecting ideas for the destination. Set the task of presenting this plan as a flow chart.
- Ask students to draw a diagram of the solution paths they explored, showing how they worked together to select the options they chose (destination, mode of transport, attractions to visit).
- When students have spent some time attempting to satisfy the ‘contradictory’ requirements, allow them to select one requirement to exclude from the list, and continue with their planning.
- Allow time for reflection at the end of the task to identify the process they followed, what worked well, and what they could have done differently.
- Ask students to reflect on and describe how the skills they learnt in doing the task could be applied to other areas of study or life.
- Before allowing students to vote on the excursions, conduct a debate with each group arguing for the excursion idea of another group.

For students working at Stage D on the cognitive skills progression, who can typically identify connections between multiple pieces of information and use systematic exploration:

- Set challenging requirements for students to address in the excursion. For example, outline specific information that needs to be collected in relation to a science or history topic. Build a connection to current learning across curriculum areas.
- Allow students to discuss, negotiate and decide on a process for planning the excursion. As part of this process, set the task of researching and selecting appropriate graphic organisers and planning and presentation tools for analysing and presenting the proposed approaches.
- Once they have tabled their ideas for an excursion, but before they have selected the one to present, change some parameters or goals to necessitate a re-planning activity. Guide students to use the same planning tools and graphic organisers to update their planning to accommodate the changes.
- Allow time for reflection to identify the process they followed, what worked well, and what they could have done differently.
- Ask students to reflect on and describe how the skills they learnt in doing the task could be applied to other areas of study or life.

Variations for Students at Different Stages on the Social Skills Progression

Some students show particular strengths in some aspects of complex tasks but are less proficient in other aspects. This is illustrated in Table 14.10, which shows, for example, that students who were working at the first stage on the cognitive skill progression could potentially be working at any one of four different stages of proficiency in the social aspects of collaborative problem solving. Teachers may need to take both of these pieces of information into consideration when planning targeted learning experiences for their students. Variations of teaching strategies, interventions or experiences for students working at different stages on the social skills progression could include those listed below.

For students working at Stage A on the social skills progression, who are typically developing their confidence to participate in collaborative tasks:

- Use small groups or pairs to allow students to become comfortable with the basic skills of collaboration. Allow students to choose partners with whom they are comfortable to work. For example, a *clock buddies*[2] strategy might be used for assigning pairs.
- Before students start the task, explicitly identify one listening skill you would like the students to demonstrate. Describe and model the listening skill. During the task, provide positive feedback when you observe students using the selected listening skill.
- Explicitly identify verbal cues in the context of discussions during the task. Where necessary, guide students through appropriate responses to these direct cues.

[2] Please refer to *Definition of Terms* at the end of the chapter.

For students working at Stage B on the social skills progression, who typically require support and scaffolding to actively participate in a collaborative task:

- Keep the group sizes small and, to promote engagement, start off with a discussion on how the task is relevant for building skills that are necessary and useful in everyday life. Link the goal or topic of the task to previous learning.
- Encourage students to 'have a go'. Discuss the consequences of avoiding risks in the context of group collaboration. Ask students to watch out for their fellow students taking a risk in participating and contributing ideas, and point it out when they do.
- Identify points during discussion where communication is not clearly understood. Ask the student to repeat what he/she said. Ask another student to explain what he/she heard. Provide opportunities for students to try alternative ways of communicating their ideas.
- Explicitly identify non-verbal cues in the context of discussions during the task. Guide students through ways to adapt a response to accommodate non-verbal cues.
- At the end of each session, allow time for reflection, and ask students to identify examples of positive behaviours and approaches displayed by themselves and others in their team.

For students working at Stage C on the social skills progression, who are typically learning to recognise their partner's role in a collaborative task:

- Ask students to each propose one idea, then re-allocate ideas to different students who must try to persuade the group to adopt the idea.
- Provide a different set of resources to individuals or sub-groups to motivate students to collaborate. For example, one group could be given access to various forms of maps, and another could have all the information relating to transport – timetables, route diagrams. They should be instructed to communicate verbally without allowing the other group visual access to their materials.
- Set goals for positive behaviours such as providing encouragement to other group members. Use a *Y chart*[3] to brainstorm how these could be recognised and the impact they could have.
- Identify opportunities for reflection and use a *freeze-frame*[4] strategy for students to discuss what is working or not working, and to identify options to proceed.
- At the end of each session, allow time for reflection and ask students to identify examples of successful strategies and positive behaviours displayed by themselves and others.

For students working at Stage D on the social skills progression, who are typically learning to share resources and information and be aware of their own and their partner's performance on a collaborative task:

---

[3] Please refer to *Definition of Terms* at the end of the chapter.

[4] Please refer to *Definition of Terms* at the end of the chapter.

- Group size could be increased to create a greater challenge in achieving a positive collaborative dynamic. In order to motivate all students to participate, the group can be given an additional task of evaluating the style and level of contribution of all group members against agreed rubrics. The students could be asked to present a pie chart that they construct jointly, to show individual contributions to the task.
- Further challenge can be provided to students by removing face to face collaboration, allowing online communication only, in the form of emails, shared documents and chats.
- Set goals for putting into practice positive group behaviours such as providing feedback to other group members to improve their contributions. Use a *Y chart*[5] to brainstorm how these behaviours could be recognised and the impact they could have.
- During the task, identify opportunities for reflection and use a *freeze-frame*[6] strategy for students to discuss what is working or not working, and to identify options to proceed.
- At the end of each session, allow time for reflection, and ask students to identify examples of successful strategies and positive behaviours displayed by themselves and others in their team.

Example: Learning in Networks – Building Intellectual Capital

Teaching and Learning Intention

This section presents ideas for teaching and learning sequences designed to develop students' skills in creating and using social media and online resources to generate new knowledge and make it accessible. Based on the work of Wilson and Scalise (2015), the skills to be developed include:

- understanding the use and purpose of tags;
- thoughtful organisation of information;
- interrogating data for meaning;
- understanding the role of social media in providing access to knowledge, sharing knowledge and creating new knowledge;
- finding and consulting experts in an online environment;
- evaluating online information;
- effective presentation of data and knowledge;
- understanding audience and cultural context; and
- creating online products for a purpose.

[5] Please refer to *Definition of Terms* at the end of the chapter.

[6] Please refer to *Definition of Terms* at the end of the chapter.

Theme: Development of a knowledge base. Students are given the task of putting together a website as a representation of a body of current knowledge on a topic. The topic can be entirely determined by the students, or selected from a list of topics as an integral part of a curriculum area. As part of the task, students are expected to use social media to access information, as well as to invite participation in a learning network, to organise information for their purpose, and to create new knowledge using multiple sources and experts.

Differentiation for Stages on the Skill Progression

The section below presents the task in different forms to suit each of the four stages on the progression of skills in building intellectual capital. The stages of proficiency progress from emerging builders, who have some knowledge of interacting online to organise information, to functional builders who have a broader perspective on how knowledge can be organised as well as an ability to collate and represent knowledge. The next stage in the progression describes proficient builders, who are able to employ social media in their purpose of seeking out relevant information to build collective knowledge. As students build their proficiency, they progress to becoming visionary builders who are able to take a leadership role in designing social media and shaping architectures for building collective knowledge.

For emergent builders:

- Provide a list of three or four topics for students to choose from for the creation of a knowledge base. Use topics related to those being covered in class, for example in science, cultural studies, or history.
- Small groups or pairs can be used for students to support one another in the use of unfamiliar technologies. There are likely to be many opportunities for peer coaching by students with particular technological skills.
- To establish preliminary skills on tagging, present a collection of tags and ask students to classify them according to different purposes such as:
    - to make a point;
    - to unify people with a common cause;
    - to collect information;
    - to promote ease of finding information;
    - to guide readers or audience; or
    - to engage interest.
- Set up different contexts for students to create their own tags and use them for the various purposes listed above, then allow students to apply this skill to the topic for the website they are developing.
- Ask students to brainstorm phrases for tags on each topic chosen by the students and then group information into tags. Keep these on display on a ‘twitter wall’ for the duration of the project. Allow students to add ‘tweets’ and reorganise tags on the wall.

For functional builders:

- In order to build awareness of presenting information in an appropriate way for an audience, discuss a variety of social media sites and set students the task of identifying the target audience.
- As a group task, ask students to identify the top three social media sites, given a set of criteria, such as age group, purpose or risk. This will involve students searching for and interpreting reviews.
- Ask students to identify the audience for whom they are building their knowledge website, and explore what that means for the way they present information.
- Set a sub-task for students to design a survey to find out about an aspect of their topic. Provide an example of a survey tool and ask students to find at least one more tool, explain which one is better for their task and provide reasons.

For proficient builders:

- Demonstrate basic use of Web 2.0 as a data visualisation tool, and allow students to present an aspect of their topic using Web 2.0.
- Present a variety of websites with contradicting viewpoints (e.g., website on moon landing, believers/sceptics regarding alien life) and conduct a discussion on how to evaluate the credibility of the websites.
- Jointly develop a set of criteria for evaluation that students can apply to the sources of information for their topics.
- Ask students to survey a population on an aspect of their topic, and post results on their websites. Discuss sources of bias that could result from the population surveyed. Show an example of a change in results when the survey population is changed.

For visionary builders:

- Start off with a session on creativity and brainstorming for students to select topics for a strategic purpose, for example, a commercially viable project, or a campaign for a social or environmental cause.
- Provide resources such as websites or tools that are of limited use. Allow inventions to be driven by necessity to promote creativity.

Variations for Students at Different Stages on Other Skills Progressions

Earlier in this chapter, adaptations or variations were suggested to cater for the range of levels of social skills in collaborative problem solving observed for a particular class of students, but the cognitive skills remained the primary focus of the learning goal. In a similar way, these suggestions can be used and adapted to suit the tasks designed to teach the building of intellectual capital and other aspects of ICT literacy skills. Once teachers become familiar with students' patterns of strengths from the profile reports (see Fig. 14.3), they can look at how to overlay multiple strategies or adjustments that address the learning needs of individual students on more than one skill domain to promote learning simultaneously on multiple progressions.

Summary

This chapter has outlined a framework for using assessments to inform the teaching of skills that were identified in the ATC21S project. The assessments are used to pinpoint a proficiency level on the empirically derived progressions for social and cognitive collaborative problem solving skills, as consumers and producers in networks, and for building social and intellectual capital in the context of ICT literacy. An understanding of each of these domains and the levels on the corresponding progressions can give teachers a starting point to understand the zone of proximal development (Vygotsky 1978) for their students – in other words, to identify which skills their students are ready to learn and should be able to develop with scaffolding from more capable others and opportunities to practise their skills. This knowledge can be developed into teaching plans and strategies to be used in the classroom, as demonstrated by the examples given in this chapter. The next step in establishing the feasibility of the approach outlined in this chapter is to evaluate the effectiveness of teaching strategies developed in this way, by checking student progress through assessments over time.

Definition of Terms

Freeze-Frame

During a group discussion there may be opportunities to re-focus the discussion or deal with conflict by instructing students to 'freeze' the discussion so that the situation can be analysed in more depth. Alternative responses or actions can be created with the benefit of time for thought. Prompts can be given such as:

'What led to this situation?'
'What were you planning to say?'
'What response do you think that would lead to?'
'What other possibilities are there?'
'How would a different question or action change the discussion?'

Y Chart

A Y Chart is a visual form of presenting ideas on how to recognise or understand the characteristics of a particular behaviour or situation. Students usually create their own Y Charts (Fig. 14.5) through brainstorming what the behaviour or situation 'looks like', 'feels like' and 'sounds like'. It helps to focus attention on observable characteristics that students can use to identify these behaviours or situations.

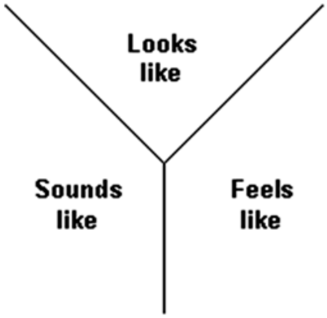


Fig. 14.5 Framework for Y-Chart brainstorming

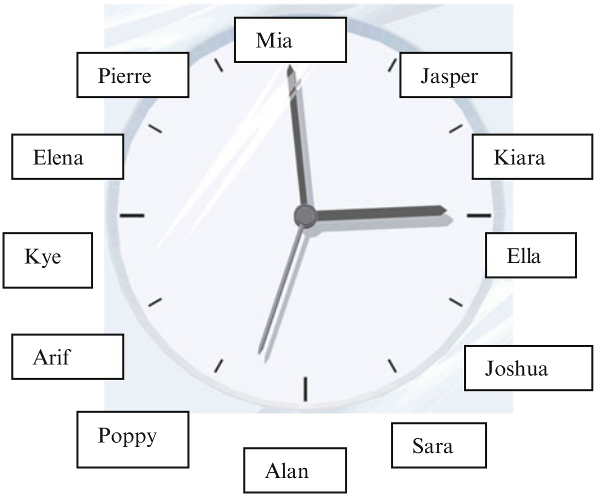


Fig. 14.6 Format for clock buddies exercise

Clock Buddies Strategy

Clock buddies (Fig. 14.6) provides a quick way of pairing students. Each student is given a clock with a space for a name beside each hour on the clock. The students are then given the task of finding a different partner for each hour on the clock and to fill in the names in the appropriate spaces, and ensure their name is in the same space on the partner's clock. The teacher can then direct the students to use their 7 o'clock buddy for a task, for example.

References

Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Raths, J., & Wittrock, M. C. (2001). L. W. Anderson & D. R. Krathwohl (Eds.), *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives* (Complete edition). New York: Longman.

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

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

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* (pp. 37–56). Dordrecht: Springer.

Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests*. (Copenhagen, Danish Institute for Educational Research), expanded edition (1980), Chicago: The University of Chicago Press.

Tomlinson, C., & McTighe, J. (2006). *Integrating differentiated instruction and understanding by design: Connecting contents and kids*. Alexandria: Association for Supervision and Curriculum Development.

Vygotsky, L. S. (1978). *Mind and society: The development of higher mental processes*. Cambridge, MA: Harvard University Press.

Wilson, M., & Scalise, K. (2015). Assessment of learning in digital networks. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills: Methods and approach* (pp. 57–81). Dordrecht: Springer.

Chapter 15
Policy Pathways for Twenty-First Century Skills

Frank Adamson and Linda Darling-Hammond

Abstract This chapter focuses on the policy environments influencing the adoption of 21st century skills in general and the results and products of the ATC21S™ project in terms of its research and assessment strategies in particular (The acronym ATC21S™ has been globally trademarked. For purposes of simplicity the acronym is presented throughout the chapter as ATC21S.). It offers a policy analysis for which information was collected from interviews with country project managers, their representatives, other participants in the ATC21S project and the international assessment community, and advisory board members from international organizations and the funding companies, supplemented by published information about national and state education systems.

Organization of the Analysis

This policy analysis first discusses how member countries have sought to incorporate 21st century skills into their education systems, independent of the ATC21S project. Over more than a decade, many countries have made progress in the integration of these problem solving, reasoning, communication, technology, and life skills into curriculum, teaching and assessment. Table 15.1 provides a comparison of approaches in different organizations and countries by mapping them to the ATC21S framework. This comparison illustrates both the significant overlap and important differences in incorporating 21st century skills in different contexts. Of course, all countries still face challenges in developing the full set of skills across all classrooms. We briefly describe the progress and these challenges.

We then discuss the piloting of current ATC21S tasks from the vantage points of educators, schools, countries and international organizations. Finally, we present feedback regarding how this type of work might be integrated into countries' curriculum and/or assessment systems, and the considerations and issues under discussion.

F. Adamson (✉) • L. Darling-Hammond
Stanford University, Stanford, CA, USA
e-mail: adamson@stanford.edu

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_15

Table 15.1 Comparison of twenty-first century skills frameworks

Assessment and teaching of twenty-first century skills (ATC21S)	European Union: key competences for life long learning (2008)	U. S. partnership for twenty-first century skills (P21)	Finnish National Curriculum 2004: cross curricular themes (C), and working methods (W)
Ways of thinking			
Creativity and Innovation;	Learning to learn	Creativity and innovation;	Human growth (C);
Critical thinking, problem solving, decision making;		Critical thinking, problem solving	Skills for thinking, learning and problem solving (W)
Learning to learn, metacognition			
Ways of working			
Communication;	Communication in the mother tongue;	Communication;	Working skills (W);
Collaboration (teamwork)	Communication in the foreign languages	Collaboration	Social skills (W);
			Active participation (W)
Tools for working			
Information literacy;	Mathematical competence and basic competencies in science and technology;	Information literacy, media literacy;	Media skills and communication (W);
ICT- literacy	Digital competence	ICT-literacy	Human technology (W);
			ICT-literacy (W)
Living in the world			
Global and local citizenship;	Cultural awareness and expression;	Flexibility and adaptability;	Cultural identity and global awareness (C);
Life and career;	Social and civic competencies;	Initiative and self-direction;	Participatory citizenship and entrepreneurship (C);
Personal & social responsibility, including cultural awareness and competence	Sense of initiative and entrepreneur-ship	Social and cross-cultural skills;	Responsibility for the environment; well-being and a sustainable future (C);
		Productivity and accountability;	Safety and traffic (C)
		Leadership and responsibility	

Source: Adapted from Ahonen (2012)

Each country has an individual policy context, but useful parallels exist across environments, and they receive attention both as future directions for member countries and as potential roadmaps for the use of ATC21S materials in the future.

Curriculum in National Contexts

Many countries and states began infusing their curriculum documents with references to 21st century skills during the 1990s. While countries generally agree about the importance of 21st century skills, some of the particular skills are easier than others to embed in curriculum and assessments. Efforts to develop critical thinking and problem solving skills are better established than research and practice around skills like creativity and innovation, ICT literacy, and collaboration, for example. Also, the skills for living in the world – such as citizenship and personal and social responsibility – do not fall squarely in the cognitive category that schools have traditionally addressed. The varied nature of 21st century skills has led to variable levels of implementation, both in a temporal sense and across different countries.

Incorporating 21st Century Skills into Curriculum

A number of the participating countries have worked for more than a decade to incorporate 21st century skills into their curricula. One of the more proactive nations in this regard is Singapore. In 1997, Singapore introduced a reform under the banner of "Thinking Schools, Learning Nation," which began a process of revising curriculum, teaching and assessment to incorporate critical thinking, problem solving, decision-making, collaboration and innovation. Curriculum documents emphasized these skills; training for new teachers began to focus on them. Some assessments were even introduced into the examination system requiring project work and investigation, as well as collaboration.

In 2010, the Singapore Ministry of Education (MOE) (2010) went further to introduce a new framework for 21st century competencies. The preparation of this framework included results from the ATC21S paper that identified and classified 21st century skills (Binkley et al. 2012). The Singaporean model also contains six core values (respect, responsibility, integrity, care, resilience, and harmony) nested within the competencies (see Fig. 15.1) (Singapore MOE 2010). Both the ATC21S categories and Singapore's framework highlight competencies that pertain both to the internal, cognitive processes of students and to their interactions with the outside world. Singapore also explicitly identifies core values as central to the learner, an approach taken by Costa Rica and Finland as well.

Finland, too, has focused on certain 21st century skills for some time, with special attention to metacognition and the development of students' abilities to "learn to learn." The Finnish national core curriculum is a lean document, reduced from

Fig. 15.1 Singapore's twenty-first century competencies and desired student outcomes (Source: http://www.moe.gov.sg/media/press/2010/03/moe-to-enhance-learning-of-21s.php)

what was once hundreds of pages of highly specific prescriptions to descriptions of a small number of skills and core concepts each year. This core curriculum guides teachers in collectively developing local curricula and assessments that encourage students to be active learners who can find, analyze and use information to solve problems in novel situations.

Finland's attention to 21st century skills deepened when the Finnish government recently renewed the discussion around a new iteration of the national curriculum. This decennial process includes writing the basic document for the national curriculum and creating corresponding legislation. The current curriculum includes some 21st century skills in the outline of goals for teachers, such as learning through peer interaction, helping students take responsibility for their learning, and helping them develop strategies for applying skills in new situations (Finnish National Board of Education 2004). The new curriculum adds to this by both including 21st century skills within the curriculum document and codifying them in specific sub-points under the education legislation. For instance, these sub-points mandate that students have the ability to collaborate and that both teachers and students increase their learning about and use of ICT (Ministry of Education and Culture 2012). While Finland remains a decentralized system with high levels of teacher autonomy, the curriculum frameworks do provide direction for teacher training, professional development and classroom practice.

The Netherlands is even more decentralized than Finland. The country does not have an official national curriculum program, due to an historical decision to equally fund public and private schools that long ago ceded curricular autonomy to schools and teachers. For this reason, decisions about whether and how to incorporate 21st century skills into curriculum and teaching occur at the school level. Some schools have worked for 20 years to introduce and improve new ways of learning. Currently, many schools seek to implement 21st century skills by having groups of students collaborate, using the internet actively, and, in some cases, engaging in international exchanges through the internet that may also include physical visits. In the "Technasia" project, which supports a network of schools, students work on technical problems with local companies to develop their cooperative problem solving.

In Australia, some states have focused on the 21st century skills for some time. For example, the ACT, Queensland and Victoria included a deepened focus on problem solving, critical thinking and communications in curriculum guidance during the 1990s and in project-based components of examinations. Queensland also developed the "Rich Tasks" project, which created assessments that included collaborative work, decision making, problem solving and metacognition. This work has informed developments in Hong Kong and Singapore.

More recently, Australia has been developing a national curriculum grounded in the school disciplines. The new national curriculum also includes a delineated set of "general capabilities." These general capabilities use different language than the ten 21st century skills presented by ATC21S. However, they do include ICT capability, critical and creative thinking, and personal and social capability, among others. A new national organization, Education Services Australia (ESA), develops curriculum materials related to the new national curriculum. It is hoped that the curricular materials will integrate the general capabilities into subject areas like mathematics, English, science, history and geography, so that teachers will be more easily able to incorporate the 21st capabilities into their teaching.

In the USA, as in Australia, the states have traditionally played the key role in education policymaking. However, initiatives to create a more centralized approach are underway. Over the last 20 years, many states and localities have developed curriculum guidance and materials incorporating elements of 21st century skills. Some states have incorporated such skills in state curriculum documents and some have integrated work on teaching for these skills in educator development programs. A private organization that includes state members, the Partnership for 21st Century Skills (P21), has assembled a website with resources for educators, policymakers and community members. P21's goal is to facilitate understanding of the role of 21st century skills from the learner to the classroom and policy.[1] Because of the historically decentralized system in the United States, and the fact that required tests do not focus on 21st century skills, the attention to such skills in national discourse and some curriculum documents does not yet translate systematically into classroom practice.

More recently, a group of 45 states has developed and adopted "Common Core State Standards" in English language arts and mathematics. These are

[1] http://www.p21.org/

aimed at creating "fewer, higher, and clearer" standards to guide curriculum, instruction and assessment. They are intended to be internationally benchmarked and to focus more intently on problem-solving and critical thinking skills, as well as on information literacy and the ability to communicate thinking and reasoning in multiple forms. New science standards under development are also expected to support a multi-state initiative. These will focus on inquiry in science education. As a companion initiative, the National Research Council recently conducted a workshop on the role of 21st century skills in science education, documented in a report covering standards, curriculum, teacher readiness and assessments (Hilton 2010). It is possible that these efforts will have some broader effect on state and local practice with respect to the cognitive skills they address. These efforts, however, do not emphasize competencies like creativity and innovation, collaboration or the life skills.

In 2003, Costa Rica enacted a *Plan of Action on Education for All*, intended to re-launch its education system (Ministerio de Educación 2002). At the center of the 2003 plan lies a view of education as a broad endeavor to develop the social, emotional and cognitive competencies that form the foundation of a 21st century skills approach. The central axis of the curriculum for this plan is a set of "values" including the daily practice of achieving a better quality personal, family and social life, and understanding human rights, health, the environment and sustainable development (Ministerio de Educación 2002). The Minister of Education, Dr. Leonardo Garnier, has validated the importance of 21st century skills in particular in a video describing the 21st century skills from ATC21S as a direction for the Costa Rican education system.[2]

As an example, within the domain of mathematics, the Ministry has introduced comprehensive curriculum reforms that focus on a student-centered approach. The new national curriculum, adopted in 2012, aims to demystify mathematics as a discipline and to motivate students by connecting their everyday experiences with mathematical principles (Garnier 2012). The curriculum focuses on five areas to achieve rigor and depth of understanding:

1. Problem solving as the main methodological strategy;
2. Contextualization as a special pedagogical component;
3. Intelligent use of digital technologies;
4. Promoting positive attitudes and beliefs about mathematics;
5. The history of mathematics.

This list incorporates some of the 21st century skills promulgated by ATC21S such as problem solving, Information literacy, and ICT literacy (Garnier 2012). The reform also proposes five attitudes that include the 21st century skill of collaboration (Garnier 2012). While Costa Rica is rolling out a 21st century skills approach beginning with particular subjects, the country has a clear national imperative to move in this direction across the entire curriculum.

[2] http://www.atc21s.cr/component/content/article/1/24-mensaje-del-dr-leonardo-garnier

Curriculum Challenges

Curriculum adoption is a national process in some countries and a state or local process in others. However, implementation is always a local process, dependent on teachers' commitments, knowledge and skills in the classroom. Some more centralized countries, such as Singapore and Costa Rica, not only have national curricula, but also have a role in the provision of materials and professional development and can use these to focus on teachers' capacities to implement pedagogies relating to 21st century skills. Australia also has begun a national curricular institute to develop curriculum materials infused with what it has termed "general capabilities," but it must incentivize states to participate and adopt these while supporting teacher-based interest in 21st century skills.

Decentralized countries such as Finland and the Netherlands face different challenges. Their pathway to adopting 21st century skills will probably not center primarily on top-down implementation of new curriculum. In Finland and the Netherlands, teacher- and school-led initiatives are the currencies of change in education. Strategies will depend more on school-initiated projects or approaches and on professional engagement of teachers through their associations or collaborative networks. In the United States, highly decentralized approaches have dominated historically; however, the Common Core standards initiative is expected to lead to a more centralized development and adoption of curriculum materials across the country. Furthermore, because of high-stakes accountability policies, assessment drives classroom practice to a substantial degree. New multi-state assessments are being developed to implement the Common Core standards, and it remains to be seen how much they will incorporate 21st century skills.

Teaching in National Contexts

Differences between centralized and decentralized countries are sometimes also related to the amount of support, professional development and guidance countries can readily make available to teachers. Centralized countries that are focused on teacher support can organize means to prepare and develop teachers more extensively. In all of these countries, there is substantial variability in classroom practice. Nonetheless, both centralized and decentralized countries have found innovative ways to support the transformation of instruction toward competency in 21st century skills.

Helping Teachers Enact 21st Century Skills

Singapore has perhaps the most supportive national environment for coherent teacher development, with only one preparer of pre-service teachers (the National Institute of Education) and substantial collaboration between NIE, the Ministry of Education and a new national Academy for supporting in-service teacher development. All have been focused on developing 21st century skills through the training of teachers to use technology, to support critical thinking and to enable collaboration, innovation and creativity. Among the innovative approaches adopted are initiatives at NIE to demonstrate to new teachers and leaders what the classroom of the future will be like and to replicate this collaborative, technology-based environment in the education of prospective educators themselves. In-service teachers are supported through action research and lesson study to develop their practice; and experimental initiatives to create school models grounded in technology-supported, inquiry-based education have been implemented and studied. These will be scaled up in ways that allow educators to support the learning of other educators.

This set of aligned efforts has begun to transform practice in Singapore, as it is supported by curriculum and assessment changes. As in every context, efforts toward change are constrained by the prior experiences of teachers (as both students and teachers), by the traditions of schooling and by the elements of the system that have not yet evolved.

Finland's teaching environment is much more decentralized. However, all teachers receive high-quality common preparation to teach. With a focus in the Finnish curriculum on "learning to learn," teachers are encouraged to cultivate students' active learning skills by posing complex problems and helping students address these problems. Teachers are taught to cultivate independence and active learning, as well as to develop metacognitive skills (Lavonen 2008). Because Finnish teachers are free to choose their methods, they define whether and how they incorporate 21st century skills in their classrooms. Many teachers also include a focus on social and collaborative skills. Nonetheless, some research shows that teachers often privilege traditional subject matter instruction over aspects of 21st century skills that go beyond disciplinary instruction. Furthermore, teachers cite these skills as the most difficult part of the curriculum. Additional opportunities for professional learning will be needed to support a pervasive adoption of all of the 21st century skills in Finnish classrooms.

Some Australian states have provided extensive professional learning opportunities for teachers, with a focus on teaching 21st century skills. New initiatives may have a more pervasive influence across the country. In 2010, for example, the Australian Institute for Teaching and School Leadership (AITSL) was launched with three primary responsibilities for working with teachers: (1) developing rigorous national professional standards; (2) fostering and driving high-quality professional development for teachers and school leaders; and (3) working collaboratively across jurisdictions and engaging with key professional bodies. This federal initiative will seek to develop stakeholder consensus in order to set new national levels and metrics for assessing teacher expertise. Ultimately, the goal is to develop a more professional, higher-status teacher workforce. With training based on the general competencies, this teacher workforce would be better suited to scaffold student learning and use assessment data to pinpoint the areas in which students require the most attention.

Challenges for Teaching

In some locations and for a few different reasons, teachers can find adopting 21st century skills difficult. First, a general dearth of example lessons with accompanying assessments leaves teachers needing to invest more time to develop their personal materials. Second, in some places, few professional development opportunities for incorporating 21st century skills exist to address the need for greater teacher knowledge about the skills. Finally, from an assessment perspective, the pressures of older test formats and differing viewpoints about the role of teachers and testing can in some cases adversely impact the teaching of 21st century skills.

Assessing 21st Century Skills

The nature of a state or national assessment system can have substantial influence over whether 21st century skills are a legitimate point of focus for teachers. The kind of influence is a function of the nature of the assessments, the stakes attached to them, and their reach. These aspects of assessment differ significantly from one nation to the next. In general, as the focus moves from assessment for accountability to assessment for learning, teachers have more space to adopt new pedagogies and foster new 21st century skills.

Influences of International Assessment

Since the development of international assessments from the *Organization of Economic Cooperation and Development* (OECD) and the *International Association for the Evaluation of Educational Achievement* (IEA), policymakers have noted the national scores in core subjects such as language arts, mathematics and science relative to other countries. Performance on these international assessments – TIMSS, PIRLS, PISA and others – has influenced the direction of national policy. For instance, the Netherlands education agenda currently focuses on core subjects in an attempt by the ministry to improve scores in these areas. On the other hand, countries like Singapore have performed well on these assessments, so stakeholders there may see changes in the system as unnecessary.

However, in 2015, PISA will begin testing collaborative problem solving (CPS), offering a political opportunity for countries to adopt approaches towards the 21st century skills that CPS involves. In addition, the IEA is now organizing an international assessment called the International Computer and Information Literacy Study (ICILS). These international assessments may focus some nations on the skills that they measure while providing models of assessment that may inform national, state, and local assessment systems.

National and State Assessment Programs

In Australia, the influence of the new international assessments will be especially direct. Julian Fraillon from the Australian Council for Education Research (ACER) is both the international director of the new IEA assessment program and the director of Australia's NAP-ICT Literacy assessment. The actual international instrument and scale of ICILS are part of Australia's national assessment work. The connection between the international and domestic assessment will allow Australia to benchmark its system's performance.

This international focus in Australia operates alongside a new national organization, the Australian Curriculum Assessment and Reporting Authority (ACARA) that is developing the national curriculum and national assessment. In addition to its national tests of language arts and mathematics, Australia has developed assessments for two skills outside of core subject areas: ICT literacy and Civics and Citizenship. Begun in 2005, the ICT Literacy assessment is a triennial national sample of Year 6 and Year 10 students "on their ability to appropriately access, manage, integrate and evaluate information, develop new understandings and communicate with others in order to participate effectively in society" (NAP 2011).

State examination systems in Australia will continue to operate at the secondary school level alongside the new national tests. The national tests are currently focused on basic literacy and numeracy skills, but there are plans to deepen their approach as the new national curriculum comes on line. The state tests vary across states in the extent to which they focus on a more traditional transmission and recall of information or cultivate deeper problem solving, collaboration, creativity and innovation. Some states and territories, such as the ACT and Queensland, require students to design and conduct inquiries and investigations, both independently and collaboratively, and encourage students to define their own problems as well as solving those posed for them.

Singapore has also signaled its interest in developing 21st century skills by infusing some forward-looking approaches in its examination system. As in the UK and some Australian states, some of the examinations include a classroom-based project component that asks students to design and manage a complex problem-solving task and to communicate about the results of the inquiry. Among these is a *Project Work* (PW) assessment, which is completed in collaborative teams as part of a compulsory interdisciplinary course for all pre-university students. The centrally-set tasks are designed to be sufficiently broad to allow students to carry out a project that they are interested in while meeting the requirements. Both product and process are assessed through a *written report,* an *oral presentation,* and a *group project file.* In carrying out the PW assessment task, students are intended to acquire self-directed inquiry skills as they propose their own topic, plan their timelines, allocate individual areas of work, interact with teammates of different abilities and personalities, and gather and evaluate primary and secondary research material. These PW processes reflect life skills and competencies, such as knowledge application, collaboration, communication and independent learning.

In the United States, accountability testing at the state level currently measures basic reading and mathematics skills through selected-response items that largely do not tap 21st century skills. Because the test results influence high-stakes decisions about students, teachers and schools, teachers typically feel they must focus on these basic skills in the formats by which they are tested at the expense of a broader array of learning objectives. However, the National Assessment of Education Progress (NAEP), a large-scale sample assessment, is broadening the types of items included in its content area tests and is launching a technology and engineering literacy assessment in 2014. This assessment will focus in part on ICT literacy as a 21st century skill in the ATC21S framework. Another possible large-scale location for assessing 21st century skills is within the new multi-state assessment consortia. The Smarter Balanced Assessment Consortium (SBAC) has requested advanced release of the ATC21S tasks, although incorporating them within the testing time windows currently under discussion seems unlikely. This consortium and another – the Partnership for Assessment of Readiness for College and Careers (PARCC) – both plan to increase assessment of problem-solving, critical thinking skills, and at least written communications skills in their tests.

Local Assessments

Finland does not conduct standardized tests developed outside of individual schools until its voluntary matriculation examination in 12th grade. Finnish education authorities periodically evaluate school-level samples of student performance, generally at the end of the 2nd and 9th grades, to inform curriculum and school investments.

All other assessments in Finland are designed and managed locally. These local assessments are, in part, guided by national curriculum documents, which indicate the kinds of assessments that teachers should develop and use to evaluate particular aspects of the curriculum within the designated subject areas. Typically, assessment guidance indicates that students should set their own learning objectives and should engage in self- and peer-assessment as well as being evaluated by the teacher. Both active learning and self-reflection are emphasized.

The ATC21S national project manager in Finland, Arto Ahonen of the University of Jyväskylä, suggests there may be value in creating non-content-based assessments of 21st century skills, because Finnish teachers tend to focus – as the national curriculum does – on the content domains they teach, more than the development of cross-cutting skills. He thinks that Finland would benefit from a generic 21st century skills assessment to promote their inclusion in Finnish classrooms. Presumably, such assessments would need to be offered to teachers for local use, since assessment is decentralized in Finland.

A wide range of interesting approaches to assessing 21st century skills occurs within schools in the United States, ranging from portfolios of research and inquiry projects, presented using oral, written, and technological tools, to more standardized assessments that have sought to incorporate 21st century skills in the evaluation of specific content. One example of the latter approach is a set of science tasks developed by the Stevens Institute of Technology and used in a group of schools for evaluating collaborative problem-solving skills. Student research groups tackle a complex problem (such as sequestering carbon from the air and the ocean), access data about different aspects of the phenomenon as though they were independent research groups working on a specialized part of the question, meet together to combine their data, and figure out a solution that requires the synthesis of data and analyses conducted across the groups. The collaboration is authentic and necessary to solving the problem. Later, they answer a series of questions to see if they understood the information and the solution. The variety of interesting approaches developed locally and in universities has not yet been taken up by state tests that determine the instruction in most schools, but it is possible that this will happen as the accountability system in the U.S. continues to evolve.

Pathways for Adopting ATC21S

A variety of pathways exists for the continuation of ATC21S in partner countries: task incorporation into assessments; use of tasks as exemplars; continued research on the learning progressions and the effects of the tasks on students; duplicating and designing similar tasks; and formative professional development for teachers. To continue along these avenues, regional connections seem possible between Singapore and Australia, the Netherlands and Finland, and Costa Rica and other Latin American countries under the auspices of the Inter-American Development Bank. Countries with similar contexts and/or goals might work together to develop more robust tasks and materials than would otherwise be possible.

The scientific work from ATC21S offers countries and research agencies significant contributions to their efforts in this arena and several members see the value in building directly upon this work: the project has developed not only sample items, but also a scale, a set of instruments, and a set of developmental progressions that can inform other work in this area (see Griffin et al. 2012).

One option that may be possible in some countries, like Australia, where there is a close connection between national goals and ATC21S tasks, would be to conduct an assessment trial that could simultaneously make people more conscious of the importance of 21st century skills and provide information to the public and policymakers on whether and how the education system is meeting the aspirations outlined in the national goals.

To achieve a larger goal of influencing classroom practice, participants believe that countries are likely to need the suite of materials developed by the project for classroom implementation. The project has made an explicit link between a defined developmental progression for students and the assessment scale mapping onto the progression. This link allows the development of teaching materials that help teachers to help students move along the progression (see Woods et al. 2015, Chap. 14).

In countries like Australia, that have been investing in the development of technology-based support systems for teaching, there is some possibility of being able to offer formative tools within an online system for assessment and curriculum materials, which might include, as well, the learning progression results.

Costa Rica has also been investing heavily in technology tools, and, with substantial ministry involvement, may look to use the tasks as exemplars for teaching and assessment that can help transform educational practice in a pervasive national effort to infuse their system with 21st century skills.

Finland and the Netherlands both view the project as more of a research and development project than a policy prospect. In Finland, the University of Jyväskylä has collaborated with another Microsoft-funded project called *Innovative Teaching and Learning* (ITL) that focuses on the teaching of 21st century skills. However the ITL 21st century skills overlap but are not derived from the same source as the ATC21S skills set identified by Binkley et al. (2012).

A number of countries envisage future research questions to pursue with ATC21S. For example, the CPS tasks use different real-world partners, and it is not yet known whether these different partners may make a difference in the outcomes. Other research possibilities include the articulation of learning progressions, the use of online assessments compared to paper-based ones, and the relative effectiveness of different measures of CPS skills or digital literacy skills.

These and other forward-looking research questions can create grist for formulating design pilots, an area of research that participants in several countries are interested in exploring.

In the U.S., it is most likely that the ATC21S materials will be useful in the near term as prototypes of what is possible in measuring collaborative problem solving and LDN-ICT literacy skills. The US NPM suggests that an online resource could provide a suite of materials to inform interested potential users. These should include: the background of the project, the definition of 21st century skills, how 21st century skills fit into curriculum and instruction, and how they may be taught and assessed.

Adoption Challenges

ICT Access

The online location of ATC21S tasks requires a certain level of ICT readiness in multiple locations. First, schools need consistent access to computers on modern equipment. In all countries, access to ICT varies between and within schools. Singapore reported a high level of ICT engagement, to the point where bandwidth to support the high level of use had become an issue. Finland and the Netherlands also have a large portion of schools and students with ICT access. Due to their size and relatively decentralized systems, the U.S. and Australia have more variable access to ICT country-wide. Costa Rica has a reasonable level of ICT access, but the rural areas are still receiving help in that domain. A planned program of one-to-one computing for rural areas may change that variability quite soon.

Second, the tasks and the servers must be capable of handling the traffic that occurs during implementation of the tasks, an issue that arose during pilot testing. Finally, different countries have different rules regarding Internet use. In the U.S., for instance, districts must have a firewall as a legal protection for children, necessitating more consideration about how to operate an assessment open to the Internet.

Task Development

From the perspective of the actual tasks, as has been noted in this volume, implementation in different countries may require translation for both language and cultural context. Finland and the Netherlands both translated the tasks, and Finland experienced some difficulty with younger students accessing Internet sites in English that they could not navigate (see Ahonen and Kankaanranta 2015; Chap. 10 and Bujanda and Campos 2015; Chap. 12). Internationally used assessments will need to plan for translation of both tasks and Internet sites that students need to access to perform the tasks.

With respect to cultural translation, Costa Rica's experience illustrates a useful example for future prototyping. Costa Rica adapted a task that includes an imaginary trip to the Arctic. For Latin Americans, the Arctic is very far away, and most children are not familiar with it. To address this, Costa Rica changed the trip location to the Antarctic and changed the subject of the research from polar bears to penguins. This example demonstrates the importance of developing tools that students in particular contexts can use with less distraction and more familiarity.

Level of Centralization

The level of centralization in countries has an effect on the ability of governments to support policy and curriculum changes that local agencies and schools will be likely to implement. Singapore has a highly centralized system with a national curriculum and national assessments. It has already built some 21st century skills into its system and has the capacity to roll out ATC21S tasks if they prove reliable measures of CPS and ICT. Costa Rica also has a somewhat centralized system with national ministry level support for the development and use of 21st century skills and a head start in certain subject areas, such as mathematics. Costa Rica would like to foster the use of 21st century skills through pre-service and in-service teacher training programs, using the tools and lessons provided by ATC21S.

Australia has new national organizations for curriculum and teaching and some national assessments, but the states still maintain a substantial amount of authority. As discussed above, the new national curriculum is expected to incorporate *general capabilities* that reflect 21st century skills. Furthermore, the national ministry is interested in using the ATC21S tasks to provide a baseline sample measure of students' capacity with 21st century skills.[3]

The U.S. has a similar federal system to Australia, with states holding the bulk of responsibility for education, but it also employs a strong national accountability system through federally-required, state-implemented high stakes testing. The importance of these assessments to many educational decisions makes changing them a challenge. However, new Common Core Standards across 45 states have led to the emergence of multi-state assessment consortia seeking to broaden the types of tasks used to assess students on these internationally benchmarked standards. As noted above, the Smarter Balanced Assessment Consortium (SBAC) has requested the CPS tasks from ATC21S as they design their assessments to understand whether the ATC21S approach is suitable for the high stakes U.S. environment.

Finland and the Netherlands have similar systems: both have Ministries of Education with little authority over classroom practice. While they offer curricular frameworks and guidelines, teachers have the freedom to decide on the content and methods of teaching. Finland has a national curriculum implemented through local level assessments, while the Netherlands leaves curriculum decisions to the schools but has some school-level accountability measures based on national assessment performance. Both countries have included 21st century skills in their curriculum frameworks for the past two decades, and Finland's newest versions are becoming even more explicit than in the past. In general, both countries are interested in the research findings offered by ATC21S so that they can further develop systems that help teachers learn how to effectively teach 21st century skills.

The different ways in which nations organize their educational and political governance certainly influence the ability of countries to incorporate the ATC21S tools into their systems. In addition to direct uses of the tools planned in some countries, ATC21S country-level NPMs have also expressed an interest in regional and/or international collaboration to continue the project research, to further explore the learning progressions and their implications for teaching, and to develop new tasks.

One strength of the project is that partnerships already exist between universities, ministries of education, research organizations, and the technology companies to facilitate both the development of materials and further research. These partnerships could demonstrate a way for new countries to become involved in the project of teaching and assessing 21st century skills using more rigorous and previously tested approaches.

[3] Note that due to a change in government in 2013, this project is no longer being pursued at a federal level in Australia. Proposals for research funding are being prepared but it is unlikely that the Australian government will support further research into C21 skills and the national curriculum.

Conclusions

Around the globe, nations have been making gradual progress in infusing 21st century skills into educational systems. Most have embedded skills of critical thinking, problem solving, decision making, communication, collaboration, and citizenship into curriculum frameworks or related documents. Some have also included skills like information literacy and ICT literacy. Less visible are expectations for the cultivation of creativity and innovation, but these, too, are beginning to become more salient.

Moving these aspirations from curriculum documents to classrooms is a more challenging task. Several policy strategies appear to be key in supporting this process:

- Developing materials that illustrate where and how these skills may be integrated into content area plans and lessons, which are the common organizers of curriculum.
- Incorporating pedagogies for teaching these skills in pre-service preparation and in ongoing learning opportunities for teachers.
- Ensuring that classroom tools are widely available for enacting these skills – including access to technologies, materials, and exemplar tasks that will allow teachers to organize and students to engage in productive activities.
- Creating assessments that can evaluate these skills and that create incentives for these abilities to be widely taught as a regular part of the curriculum.
- Developing an understanding of how these capacities may develop over time – with opportunity, scaffolding, and instruction – so that teachers can envision how to organize supports for learning in these complex domains.

Countries participating in this project have undertaken varieties of these strategies, and most have had success in some domains while looking to develop their practice in others. The advent of international assessments of collaborative problem solving (OECD) and ICT literacy (IEA) will spur further interest and opportunity for research and development.

ATC21S provides critical support for these efforts by offering exemplars for defining and assessing key skills such as collaborative problem solving and ICT literacy. These can inform curriculum and assessment development in both national/state systems and local schools. ATC21S consists of more than model assessment tasks, however. A critically important development provided by the project is the creation of defined developmental progressions for students as they acquire these skills, and assessment scales which map onto the progressions. This will allow the development of curriculum and teaching materials that teachers can use to help their students reach the next step in the progression with a fairly high level of confidence (see Woods et al. 2015; Chap. 14).

(continued)

Next steps in building upon this promising beginning should include the following:

- Investment in developing a body of tasks, adapted for different cultural contexts, that can be used to evaluate student learning along the progressions and to support research on how task design features affect performance.
- Conduct of wider assessment trials that can provide benchmarking information about performance of students as well as diagnostic information about the behavior of tasks and the learning and needs of both teachers and students.
- Integration of research on specific classroom practices for developing 21st century skills, such as that underway in the *Innovative Teaching and Learning* project, with the use of ATC21S assessment tasks and tools.
- Development of teaching materials that embed the tasks as assignments in formative learning opportunities linked to the developmental progressions.
- Training of educators to develop and score tasks, and to contribute to the body of teaching materials, so that they deeply understand the underlying theories of learning and performance that can help them deepen their instruction.
- Creation of online platforms that offer a suite of educative materials, including an explication of the goals and meaning of specific 21st century skills, links to relevant national/state curriculum standards or frameworks, discussion of the learning progressions, curriculum and teaching materials, and embedded exemplar tasks with information about scales and scoring.
- Inclusion of expectations for teaching 21st century skills in teacher education curricula, and enlistment of teacher educators in creating practical models (courses, clinical experiences, and materials) for developing pedagogies to teach such skills in pre- and in-service preparation.

These steps will be enhanced by policy efforts to extend the availability of technology tools in classrooms and to include such skills as collaborative problem solving and ICT literacy in formal assessment systems, to signal the importance of moving affirmatively into the 21st century.

References

Ahonen, A. K., & Kankaanranta, M. (2015). Introducing assessment tools for 21st century skills in Finland. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills: Methods and approach* (pp. 213–226). Dordrecht: Springer.

Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., & Rumble, M. (2012). Defining twenty-first century skills. In *Assessment and teaching of 21st century skills*. Dordrecht: Springer.

Bujanda, M. E., & Campos, E. (2015). The adaptation and contextualization of ATC21S™ by Costa Rica. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills: Methods and approach* (pp. 245–256). Dordrecht: Springer.

Finnish National Board of Education. (2004). *National core curriculum for basic education 2004*.

Garnier, L. (2012). *Nuevos Programas de Matemática: Una Reforma Indispensable.* Ministerio de Educación Pública de Costa Rica. San José: Costa Rica.

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

Hilton, M. (2010). *Exploring the intersection of science education and 21C skills: A workshop summary*. Washington, DC: National Research Council.

Lavonen, J. (2008). *Reasons behind Finnish students' success in the PISA scientific literacy assessment*. University of Helsinki, Finland. Retrieved on 8 Sept 2008, from http://www.oph.fi/info/finlandinpisastudies/conference2008/science_results_and_reasons.pdf

Ministerio de Educación Pública de Costa Rica. (2002). *Plan de Acción de la Educación para Todos 2003–2015*. Costa Rica: Author. Retrieved from: http://www.oei.es/quipu/costarica/PLANDEACCION2003_2015.pdf

Ministry of Education and Culture. (2012). *Education and research 2011–2016: A development plan*. Reports of the Ministry of Education and Culture, Finland.

National Assessment Program (Australia). 2011. *ICT literacy*. Retrieved from: http://www.nap.edu.au/NAP_Sample_Assessments/About_each_domain/ICT_Literacy/index.html

Singapore Ministry of Education. (2010). *MOE to enhance learning of 21st century competencies and strengthen art, music and physical education*. Press release retrieved from http://www.moe.gov.sg/media/press/2010/03/moe-to-enhance-learning-of-21s.php

Woods, K., Mountain, R., & Griffin, P. (2015). Linking developmental progressions to teaching. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills: Methods and approach* (pp. 267–292). Dordrecht: Springer.