

# Draft White Paper 1

## Defining 21st century skills

Marilyn Binkley, Ola Erstad, Joan Herman, Senta Raizen, Martin Ripley  
with Mike Rumble

The Assessment and Teaching of 21st Century Skills project was created by Cisco, Intel and Microsoft and launched at the Learning and Technology World Forum 2009 in London. During 2009, the project operated with five Working Groups, each of which produced a White Paper. These papers will be fully edited into a volume that will be published electronically on the project website ([www.atc21s.org](http://www.atc21s.org)). Print publication is also being considered.

As a report to the Learning and Technology World Forum 2010 in London, final drafts of the papers are collected together in this set and posted on the project website for Forum participants and others who can freely access them on the website. These drafts are not for formal citation. All persons registered on the project website for updates will be advised when the final publication has been posted on the site.

January 2010



# Contents

## Table of Contents

Contents .....	iii
Table of Contents .....	iii
List of tables .....	iv
List of figures .....	iv
List of boxes .....	iv
Abstract .....	v
Defining 21st century skills .....	1
The role of standards and assessment in promoting learning .....	2
The importance of standards that promote learning .....	2
Assessment systems that promote learning .....	3
The nature of quality assessment systems .....	4
Learning-based assessment systems .....	4
Improving the quality of assessment systems .....	5
Principles for 21st century standards and assessments .....	6
Using technology to transform assessment and learning .....	7
Assessment priorities enabled by information and communication technology .....	7
The migratory strategy with ICT .....	9
The transformational strategy with ICT .....	10
Arriving at a model 21st century skills framework and assessment .....	13
Existing 21st century skills frameworks .....	13
<i>Personal learning &amp; thinking skills – The national curriculum for England</i> .....	14
The KSAVE model .....	15
Ways of thinking .....	16
Creativity and Innovation .....	16
eSCAPE .....	17
Critical thinking, problem solving and decision making .....	17
Primum .....	19
World Class Tests .....	19
The VPA Project .....	19
Learning to learn and metacognition .....	20
eVIVA .....	20
Cascade .....	21
Ways of working .....	21
Communication .....	21
Collaboration and teamwork .....	23
Tools for working .....	24
Information literacy .....	27
ICT literacy .....	28
Living in the world .....	29
Citizenship, global and local .....	30
Life and career .....	31
Personal and social responsibility .....	33
Challenges .....	33
Using models of skill development based on cognitive research .....	34

Transforming psychometrics to deal with new kinds of assessments .....	34
Making students' thinking visible.....	34
Interpreting assisted performance .....	34
Assessing 21st century skills in traditional subjects .....	35
Accounting for new modes of communication .....	35
Including collaboration and teamwork .....	35
Including local and global citizenship.....	35
Ensuring validity and accessibility.....	35
Considering cost and feasibility .....	36
White Paper 1 Annex: Expanded descriptions of assessment tasks .....	37
UK ICT Test for 14-year-old students .....	37
eSCAPE .....	41
Primum® Computer-based Case Simulations (CCS) for licensing doctors.....	47
World Class Tests .....	55
Virtual Performance Assessment.....	59
References and further reading .....	63

## List of tables

Table 1: Sources of documents on 21st century skills .....	14
Table 2: Ways of thinking - creativity and innovation .....	16
Table 3: Ways of thinking - critical thinking, problem solving, decision making.....	18
Table 4: Ways of thinking - learning to learn, metacognition .....	20
Table 5: Ways of working - communication .....	22
Table 6: Ways of working - collaboration, teamwork.....	23
Table 7: Tools for working - information literacy .....	25
Table 8: Tools for working - ICT literacy.....	27
Table 9: Living in the world - citizenship, local and global .....	30
Table 10: Living in the world - life and career .....	32
Table 11: Living in the world - personal and social responsibility .....	33

## List of figures

Figure 1: Integrated assessment system .....	4
Figure 2: The dimensions of e-assessment innovations .....	8

## List of boxes

Box 1: Innovative UK assessment of ICT skills of 14-year-olds.....	11
Box 2: Educational Testing Service framework for ICT literacy .....	28

## **Abstract**

This paper synthesizes research on the role of standards and assessment in promoting learning, describes the nature of assessment systems that can support changes in practice, illustrates the use of technology to transform assessment systems and learning, and proposes a model for assessing 21st century skills.

Large-scale assessments should be only part of any system to support student learning. Assessments at each level represent a significant opportunity to signal the important learning goals that are targeted by the broader system as well as to provide valuable, actionable data for policy and practice. Moreover, they can model next generation assessments that can support learning. To do so assessments should a) be aligned with the development of significant 21st century goals, b) be adaptable and responsive to new developments, c) be largely performance-based, d) add value for teaching and learning by providing information that can be acted on by students, teachers, and administrators, e) meet the general criteria for good assessments, (i.e. be fair, technically sound; valid for purpose, and part of a comprehensive and well-aligned system of assessments at all levels of education)

The model for assessments of 21st century skills, based on an analysis of curriculum and assessment frameworks for 21st century skills developed around the world, identifies ten important skills in four broad categories. The paper provides measureable descriptions of the skills, considering knowledge, skills, and attitudes, values and ethics (advanced as the KSAVE framework).

The paper concludes with a discussion of challenges to be addressed in developing an assessment system that supports learning using, for example, research-based models of skill development and assessments that make students' thinking visible to establish their strengths and weaknesses and help shape future learning choices.



# Defining 21st century skills

Marilyn Binkley, Ola Erstad, Joan Herman, Senta Raizen, Martin Ripley  
with Mike Rumble

There has been a significant shift in advanced economies from manufacturing to emphasizing information and knowledge services. Knowledge itself is growing ever more specialized and expanding exponentially. Information and communication technology is transforming the nature of how work is conducted and the meaning of social relationships. Decentralized decision-making, information sharing, teamwork and innovation are key in today's enterprises. No longer can students look forward to middle class success in the conduct of manual labor or use of routine skills - work that can be accomplished by machines. Rather, whether a technician or a professional person, success lies in being able to communicate, share, and use information to solve complex problems, in being able to adapt and innovate in response to new demands and changing circumstances, in being able to marshal and expand the power of technology to create new knowledge and expand human capacity and productivity.

Research during the last decade has shown how new social practices evolve due to increased use of new digital technologies, especially among young people (Buckingham & Willett, 2006). Such practices create re-conceptions of key skills, not defined from a systems level, but from the everyday lives of people in our societies. One example is research done on computer games and online communities (Gee, 2007), where problem solving is defined as a key component of such practices. Such experiences of problem solving among young people need to inform us in the way we design assessment tasks and define key skills. Hence, new standards for what students should be able to do must replace the basic skills and knowledge expectations of the past. To meet this challenge schools must be transformed in ways that will enable students to acquire the sophisticated thinking, flexible problem solving, collaboration and communication skills they will need to be successful in work and life. New conceptions of educational standards and assessment, the subject of this white paper, are a key strategy for accomplishing the necessary transformation. Such standards and assessment can both focus attention on necessary capacities and provide data to leverage and evaluate system change. Technology too serves as both a driver and lever for the transformation.

In the sections that follow, we

- Synthesize research on the role of standards and assessment in promoting learning,
- Describe the nature of assessment systems that can support changes in practice and use these to develop guiding principles for the design of the next generation assessments,
- Illustrate the use of technology to transform assessment systems and learning, and
- Propose a MODEL for assessing 21st century skills.

Our intent is to learn from the past as we prepare for new futures in educational standards and assessment. While we provide a list of 21st century skills based on our analysis of twelve relevant frameworks drawn from a number of countries (see p.13), these serve as an example of how to think about assessing 21st century skills. We expect that educators, as they consider our model, may need to make adaptations that fit their own contexts as they design assessments appropriate for their own schools and students.

We have organized the ten skills we have identified into four groupings as follows:

## **Ways of Thinking**

1. Creativity and innovation
2. Critical thinking, problem solving, decision making
3. Learning to learn, Metacognition

### **Ways of Working**

4. Communication
5. Collaboration (teamwork)

### **Tools for Working**

6. Information literacy
7. ICT literacy

### **Living in the World**

8. Citizenship – local and global
9. Life and career
10. Personal & social responsibility – including cultural awareness and competence

## **The role of standards and assessment in promoting learning**

### **The importance of standards that promote learning**

Worldwide research has established the significant role that curriculum standards and assessment can play in molding new expectations for learning. Although the terminology of standards-led reform initially may have been associated with accountability and improvement initiatives in the United States (e.g., New Standards Project, 1998, No Child Left Behind ACT, 2001 ), the approach has widespread currency in educational systems as far spread as England, Germany, Norway, Singapore and Australia to name just a few. The basic ideas followed by these accountability and school improvement systems have rested on three principles:

- Be clear about expectations by establishing standards;
- Develop high visibility (sometimes referred to as high stakes) assessments based on the standards; and
- Use the assessments to communicate what is expected, to hold relevant stakeholders accountable and to publish data to inform decisions

Such standards-based assessments provide empirical evidence for judging performance and can serve a variety of decision-making purposes (accountability, selection, placement, evaluation, diagnosis, or improvement) but the very existence of the assessments and the attention they engender carry important social, motivational and political consequences.

Researchers around the globe studying such assessments have found fairly uniform effects. This is documented by a number of examples: studies of state accountability assessments in more than a dozen states in the United States, of A- or GCSE or Key Stage Exams in England, and of language and higher education admissions testing programs in countries such as Australia, Central and Eastern Europe, China, Hong Kong, Israel, Japan, New Zealand and Sri Lanka document that. In summary:

- *Assessments signal priorities for curriculum and instruction; high visibility tests serve to focus the content of instruction.* School administrators and teachers pay attention to what is tested, analyze test results, and adapt curriculum and teaching accordingly.
- *Teachers tend to model the pedagogical approach reflected on high-visibility tests.* When high visibility assessments are composed of multiple-choice items, teachers tend to rely heavily on multiple-choice worksheets in their classroom instruction and emphasize lower level cognitive skills. However, when the assessments use extended writing and/or performance assessments, teachers incorporate similar activities in their classroom practice.
- *Curriculum developers, particularly commercial interests, respond to important tests by modifying existing textbooks and other instructional materials and/or developing and marketing new ones to address test expectations.* These products in turn may become primary resources that influence practice and also influence teachers' understandings of test expectations.



At the same time research documents effects that can propel productive changes in practice, so too it shows the potential for substantial negative consequences.

- *Schools and teachers tend to focus on what is tested rather than underlying standards or learning goals*, and to ignore what is not tested. Both the broader domain of the tested disciplines and important subjects that are not tested may get short shrift. In the United States, England and other countries, tests tend to give relatively little attention to complex thinking and problem solving and tend to focus on lower levels of learning, which can lead to similar emphases in classroom practice.
- *Focusing on the test, rather than underlying learning, may encourage a one-time performance orientation and transmission-type teaching*. When doing well on the test, rather than learning, becomes the goal, schools may unwittingly promote a performance orientation in students, which in turn can work against students' engagement and persistence in learning, metacognition, and self regulation. Especially for high visibility multiple-choice tests, teachers may concentrate on helping students acquire specific content, rather than helping students build conceptual understandings and problem-solving capabilities.
- *Instructional/teaching time is diverted to specific test preparation activities*. Schools provide students with practice on the specific types of tasks and formats that are expected on the test, through commercial test preparation packages, special classes and homework. Such activities aim specifically to help students do well on the test, rather than promoting students' learning, and depending on the school and the pressure to improve test scores, can divert weeks or more of instructional time.

These consequences and caveats underscore an important challenge in using assessments to promote 21st century skills. The research clearly shows that whatever is measured matters; that educators tend to model and mimic in their curriculum and instruction the content and format of high visibility assessments and to use a significant amount of classroom time for special test preparation activities. In some countries, however, testing has become dominated by routine, and highly predictable items, which are also often short and highly scaffolded, thus reducing the expectation that students should apply knowledge, skills and broader capabilities demanded by today's world. For example, analyses of annual state, standards-based tests in the United States show a preponderance of items addressing lower level cognitive demand to the detriment of complex thinking and problem-solving applications (see Webb, 1999). Other countries provide more promising examples. For instance, end of secondary school/University access examinations such as the Baccalaureat, the Matura, Abitur, etc. probe in depth the content and skills that students are expected to acquire and call on students to demonstrate their knowledge and skills in a wide variety of oral and written formats and project-based work. And in the Nordic countries there is a tradition of integrating project work into the curriculum promoting more locally adapted and general standards for assessment. Such examples involve students in important, authentic performances. Even so, the assessment standards for these exams have not yet been fully updated to reflect the demands of an information and innovation age, nor do they take advantage of 21st century technology. Just as students need to be literate in new media and be able to harness their power, so too technology can open up new, cost-effective possibilities for the design and use of a new generation of assessments.

### **Assessment systems that promote learning**

The contrast between US-type accountability exams and promising, secondary and university access examinations also is noteworthy in that the latter are embedded in coursework rather than external to it, where they can become an integral part of the teaching and learning process. The exams establish meaningful goals on which course assignments and assessments can be built and are used regularly to assess and respond to student progress. Research, in fact, shows the powerful effect that on-going assessment, so called formative assessment, has on student learning, particularly for low ability students (Black and Wiliam, 1998; OECD, 2005).



**Figure 1: Integrated assessment system**

The use of assessment information is key to the idea: to be considered formative, assessment evidence must be *acted upon* to inform subsequent instruction. Rather than focusing backward on what has been learned, formative assessment helps to chart the learning road forward, by identifying and providing information to fill any gaps between the learners' current status and goals for learning. Moreover, more than solely a source of evidence that informs subsequent teaching and learning, carefully crafted formative assessments can directly support the learning process by incorporating principles of learning and cognition (Herman & Baker, 2009; Bennett & Gitomer, 2009). For example, by asking students to make public their thinking, formative probes can provide scaffolding that helps students confront their misconceptions, refine and deepen their understandings and move to more sophisticated levels of expertise (refs). By asking students for explanations and providing practice over multiple and authentic contexts, assessment tasks can help students to connect new knowledge to their existing structures and build transfer capability. By making learning goals explicit and involving students in self-assessment, formative assessment also can promote students as agents in their own learning, increasing student motivation, autonomy, and metacognition, as well as learning. Such characteristics similarly can be incorporated into accountability assessments to increase their learning value.

## **The nature of quality assessment systems**

### **Learning-based assessment systems**

Assessment design and development must bring together the rich, existing research base on student learning and how it develops with state of the art psychometric theory to produce a new generation of assessments. As a prominent panel in the United States put it:

Every assessment ... rests on three pillars: a model of how students represent knowledge and develop competence in a subject matter domain; tasks or situations that allow one to observe students' performance; and an interpretation method for drawing inferences from the performance evidence thus obtained (Pellegrino et al., 2001, p. 2).

Adapting this general model, Figure 1 is intended to communicate that quality assessment starts (and ends) with clearly specified and meaningful goals for student learning (see also Baker, 2007; Forster & Masters, 2004; Wilson & Sloane, 2000). The assessment task vertex signals that any

learning-based assessment must elicit responses that can reveal the quality of student understandings and/or where students are relative to the knowledge and skills that comprise intended learning goals. The interpretation link reinforces the idea that responses from assessment tasks must be specially analyzed and synthesized in ways that reveal and support valid inferences that connect to intended uses of the assessment, while the use vertex highlights that results must be used for student learning relative to initial goals. Assessment quality then resides in the nature of the relationships between and among all three vertices and their connections -- in the relationship between learning goals and tasks used to assess their development, in how well the analysis and scoring schemes capture important dimensions of intended understandings and skills and how well they support use are used to improve learning. Inherent here too are the more traditional dimensions of validity, accuracy and fairness of interpretations of student learning and--particularly for external and higher stakes tests--evidence that interpretations and inferences are justified (see White Paper 2).

As Figure 1 shows, there are multiple levels for which data may be gathered and used for various decision-making purposes, from on-going data to inform and enrich classroom teaching and learning (see White Paper 4) to periodic data to support policy and practical decision-making at higher levels of the educational system – e.g., school, district, province or state, national. Importantly, large scale international, national and/or state or provincial assessments, for example, may provide policymakers a general barometer for judging and responding to schools' progress in promoting student learning, for allocating resources, identifying locales that need help, etc. Schools and teachers may use the same data to evaluate their programs, refine their curricula, frame improvement plans, and/or identify individual students who need special attention. But to fuel on-going decisions to optimize teaching and learning, teachers need a more continuous flow of data. Figure 1 implies a system of assessments, grounded in a common, well-specified set of learning goals, that is purposively designed to satisfy the decision-making needs of all actors within and across the educational enterprise. Such a system needs to be aligned with the 21st century skills that will enable students' future success. Large-scale assessments can serve an important function in communicating and signaling what these skills are as well as provide important models of how they can be assessed.

### Improving the quality of assessment systems

This system perspective also requires a different vantage point for considering assessment quality. Rather than focusing only on a single test, we need to consider the quality of the system for providing valid evidence to support the varied decision-making needs at multiple levels of the educational system. Balanced assessment seems an over-riding criterion (Bell, Burkhardt & Swan, 1992). Pellegrino et al. (2001), for example, argued for the development of balanced assessment systems to serve both accountability and policy purposes as well as those of improving classroom teaching and learning. A balanced system, in their view, incorporates three critical principles: *coherence, comprehensiveness, and continuity*.

- A *coherent* assessment system is built on a well-structured conceptual base—an expected learning progression, which serves as the foundation both for large scale and classroom assessments. That foundation should be consistent and complementary both across administrative or bureaucratic levels of the educational system and across grades.
- A *comprehensive* assessment system uses a range of assessment methods to ensure adequate measurement of intended constructs and measures of different grain size to serve decision-making needs at different levels of the education system. Inherently, a comprehensive assessment system also is useful in providing productive feedback, at appropriate levels of detail, to fuel accountability and improvement decisions at multiple levels.
- *Continuity* captures the principle that assessment at all levels is conceived as part of a continuous stream of evidence that tracks the progress of both individual students and

educational programs over time. This can only be possible when there is consistency in the definition of the constructs across time, e.g., from the beginning of the year to the end and across grades.

While inherent in the above formulation, *fairness* also is a fundamental principle for assessment systems. All assessments should be designed to enable the broadest possible population of students to show what they know, without being unfairly hampered by individual characteristics that are irrelevant to what is being assessed. For example, students who are not proficient in the language of the test and test items may well find it difficult to show their mathematics capability; students from one culture may lack the background knowledge to deal with a reading passage about a context with which they are unfamiliar. Disabled or very low ability students may be below the learning threshold on which a test is based. A fair system of assessment offers accommodations for students who may need them and is sensitive to the range of student abilities and developmental levels likely in the assessed population

### **Principles for 21st century standards and assessments**

While it should be clear that large scale state, national, regional, or international assessments should be conceived as only part of any system to support student learning, assessments at each level represent a significant opportunity to signal the important learning goals that should be the target of the broader system as well as to provide valuable, actionable data for policy and practice. Moreover, carefully crafted, they can model next generation assessments that through design and use can support learning. To do so, our review to this point suggests that 21st century standards and assessments should:

- *Be aligned with the development of significant, 21st century goals:* Assessments that support learning must explicitly communicate the nature of expected learning. Standards and assessments must fully specify the rich range of 21st knowledge and skills students are expected to understand and apply. In addition, the standards and assessments should ideally represent how that knowledge and set of skills is expected to develop from novice to expert performance.
- *Incorporate adaptability and unpredictability:* One hallmark of 21st century demands is the need to adapt to evolving circumstances and to make decisions and take action in situations where prior actions may stimulate unpredictable reactions that in turn influence subsequent strategies and options. Dealing with such uncertainty is essential, but represents a new challenge for curriculum and assessment
- *Be largely performance-based.* The crux of 21st century skills is the need to integrate, synthesize and creatively apply content knowledge in novel situations. Consequently, 21st century assessments must systematically ask students to apply content knowledge to critical thinking, problem solving, and analytical tasks throughout their education, so that we can help them hone this ability and come to understand that successful learning is as much about the process as it is about facts and figures.
- *Add value for teaching and learning.* The process of responding to assessments can enhance student learning if assessment tasks are well crafted to incorporate principles of learning and cognition. For example, assessment tasks can incorporate transfer and authentic applications, and can provide opportunities for students to organize and deepen their understanding through explanation and use of multiple representations.
- *Make students' thinking visible.* The assessments should provide a window into students' understandings and the conceptual strategies a student uses to solve a problem. Further by making students' thinking visible, assessments thus provide a model for quality practice.

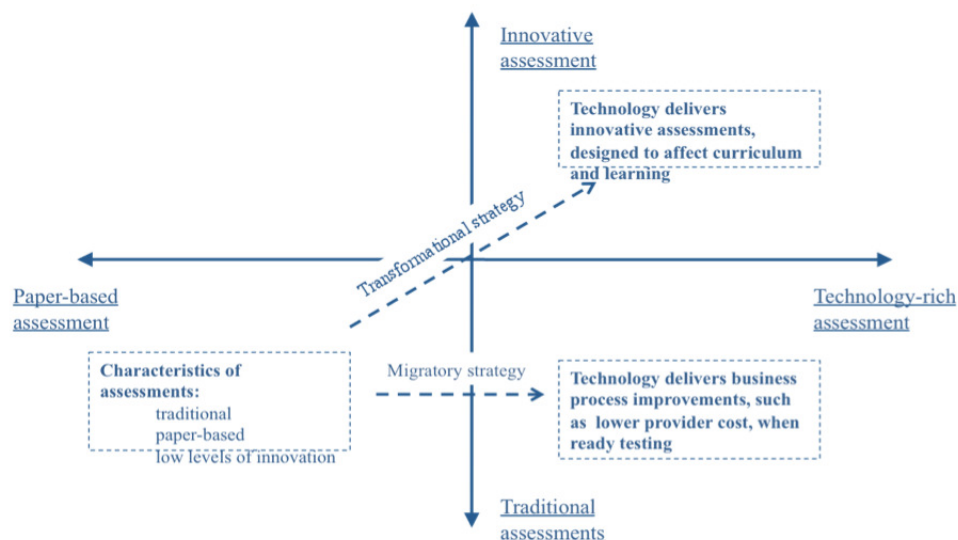
- *Be fair.* Fair assessments enable all students to show what they know and provide accommodations for students who otherwise would have difficulty accessing and responding to test items for reasons other than the target of the assessment.
- *Be technically sound.* Assessment data must provide accurate and reliable information for the decision-making purposes for which they are intended to be used. If there is an absence of reasonable precision of measurement, inferences from results and decisions based on them may well be faulty. The requirement for precision relative to intended purposes means both that intended uses and users must be clearly specified and evidence of technical quality must be established for each intended purpose. Establishing evidence of quality for innovative approaches to assessing 21st century skills may well require new psychometric approaches.
- *Valid for purpose.* To the extent an assessment is intended to serve as an indicator of schools' success in helping students acquire 21st century skills and skills, test results must be both instructionally sensitive and generalizable. That is, instructionally sensitive tests are influenced by the quality of instruction: students who receive high quality instruction should out-perform those who do not; the alternative is that students' basic ability or general intelligence, which are not under a school's control, are the reason for performance. A generalizable result transfers to other real life applications.
- *Generate information that can be acted upon and provides productive and usable feedback for all intended users.* Teachers need to be able to understand what the assessment reveals about students' thinking. And school administrators, policymakers, and teachers need to be able to use this assessment information to determine how to create better opportunities for student learning.
- *Provide productive and usable feedback for all intended users.* It seems axiomatic that if stakeholders, for example, teachers, administrators, students, parents, and the public at large are expected to use the results of an assessment, they must have access to reports that are accurate, understandable and usable.
- *Build capacity for educators and students.* Feed back from assessments can help students, teachers, administrators and other providers to understand the nature of student performance and the learning issues that may be impeding progress. Teachers and students should be able to learn from the process.
- *Be part of a comprehensive and well-aligned system of assessments designed to support the improvement of learning at all levels of the educational hierarchy.*

## **Using technology to transform assessment and learning**

The following sections of this paper mainly address large-scale assessments. White Paper 4 deals more explicitly with classroom assessments.

### **Assessment priorities enabled by information and communication technology**

In this section we draw attention to three areas where ICT has greatly increased the potential for assessing 21st century skills. ICT can be thought of as a tool for traditional assessments but also as presenting new possibilities for assessing skills formerly difficult to measure. ICT also develops new skills of importance for the 21st century. As much as we need to specify the skills needed, we also need to specify approaches that might measure the extent to which students have acquired them. During the last decade several initiatives have explored how ICT might be used for assessment purposes in different ways in different subject domains. The discussion below is based on a review of relevant research in this area.



**Figure 2: The dimensions of e-assessment innovations**

Although assessment in education is a substantial research field, it has only been during the last decade that ICT-based assessment has been growing as a research field (McFarlane, 2003). This is partly due to an increase in developments of the ICT infrastructure in schools with expanded access to hardware, software, and broadband Internet connections for both students and teachers. Existing research has examined both the impact of ICT on traditional assessment methods and how ICT raises new issues of assessment and skills. For example, as part of the Second International Technology in Education Study (Kozma, 2003), innovative ICT-supported pedagogical practices were analyzed. In several countries some of these practices showed a shift towards more use of formative ways of assessment when ICT was introduced (Voogt & Pelgrum, 2003). However, in most practices often old and new assessment methods coexisted, because schools had to relate to national standards and systems over which they have no control, while at the same time they were developing alternative assessment methods for their own purposes.

The use of the term e-assessment has gained acceptance in recent years. Advocates of e-assessment frequently point to the efficiency benefits and gains that can be realized. These benefits might have to do with the costs of test production, the ability to re-use items extensively, to create power- and adaptive-tests, or to build system improvements, such as test administration systems which are able to provide tests whenever students want to take them. However, in the report *Effective practice with e-assessment* (Whitelock, Road & Ripley, 2007) the writers conclude that e-assessment is "much more than just an alternative way of doing what we already do". Through evidence and case studies, the report provides examples of e-assessment widening the range of skills and knowledge being assessed, providing unprecedented diagnostic information, and supporting personalisation (Ripley, 2007). Thus, we argue that e-assessment has the potential of using technology to support educational innovation and the development of 21st century skills, such as complex problem solving, communication, team work, and creativity and innovation.

Figure 2 provides a representation of the contrast between the two drivers: the business efficiency gains versus the educational transformation gains. The lower-left quadrant represents traditional assessments, typically paper-based and similar year-on-year. Most school and college-based assessments are of this type. Moving from the lower-left to the lower right quadrant represents a migratory strategy in which paper-based assessments are migrated to a screen-based environment. Delivery is more efficient but assessments are qualitatively unchanged. In contrast, moving to the upper-right quadrant represents a transformational strategy with which technology is used to

support innovative assessment designed to influence (or minimally to reflect) innovation in curriculum design and learning.

### **The migratory strategy with ICT**

Conceptions of 21st century skills include some familiar skills that have been central in school learning for many years, such as information processing, reasoning, enquiry, critical thinking and problem solving. The question is: To what extent does ICT enhance or change these skills and their measurement? Indeed, during the last decade most of the research on the use of ICT for assessment has dealt with the improvement of assessment of traditional skills--improvement in the sense that ICT has potentials for large-scale delivery of tests and scoring procedures, easily giving the learner accessible feedback on performances. For example, many multiple choice tests within different subject domains are now online. The focus is then on traditional testing of reasoning skills and information processing among students, on memorization and reproduction of facts and information. Using online tests will make this more cost effective and less time consuming. However, there are several concerns raised about assessment of traditional skills in an online setting, especially regarding security, cheating, validity and reliability.

Many countries and states have adopted a “dual” program of both computer-based and paper-and-pencil tests. Raikes & Harding (2003) mention examples of such dual programs in some states in the US where students switch between answering computer-based and paper-and-pencil tests. The authors argue that assessments need to be fair to students regardless of their schools’ technological capabilities and the need to avoid sudden discontinuities so that performance can be compared over time. This may require a transitional period during which computer and paper versions of conventional external examinations run in parallel. They sketch some of the issues (costs, equivalence of test forms, security, diversity of school cultures and environments, and technical reliability) that must be solved before conventional examinations can be computerized. In a meta-evaluation of initiatives in different states in the US, Bennett (2002) shows that the majority of these states have begun the transition from paper-and-pencil tests to computer-based testing with simple assessment tasks. However, he concludes, “If all we do is put multiple-choice tests on computer, we will not have done enough to align assessment with how technology is coming to be used for classroom instruction” (pp. 14–15).

Recent developments in assessment practices can be seen as a more direct response to the potential of ICT for assessment. An example of such developments is the effort to use computers in standardized national exams in the Netherlands, going beyond simple multiple choice tests. The domain for the assessment is science, where exams contain 40% physics assignments which have to be solved with computer tools like modelling, data video, data processing and automated control technique (Boeijen & Uijlings, 2004).

Several studies comparing specific paper-and-pencil testing with computer-based testing have described the latter as highly problematic, especially concerning issues of test validity (Russell et al., 2003). Findings from these studies, however, show little difference in student performance (Poggio et al., 2005), even though there are indications of enough differences in performance at the individual question level to warrant further investigation (Johnson & Green, 2004). There are differences in prior computer experience among students and items from different content areas can be presented and performed on the computer in many different ways, which have different impacts on the validity of test scores (Russell et al., 2003). While some studies provide evidence of score equivalence across the two modes, computerized assessments tend to be more difficult than paper-and-pencil versions of the same test. Pommerich (2004) concludes that the more difficult it is to present a paper-and-pencil test on a computer, the greater the likelihood of mode effects to occur. Previous literature (Russell, 1999; Pommerich, 2004) seems to indicate that mode differences typically result from the extent to which the presentation of the test and the process of taking the test differ across modes, rather than from differences in content. This may imply a need to try to minimize differences between modes. A major concern is whether computer-based testing

meets the needs of all students equally and whether some are advantaged while others are disadvantaged by the methodology.

In a recent special issue of the *British Journal of Education Technology* focusing on e-assessment several studies are presented where students' traditional skills are assessed in different ways (Williams & Wong, 2009; Draper, 2009; Shephard, 2009.)

The introduction of ICT has further developed an interest in formative ways of monitoring and assessing student progress. The handling of files and the possibility of using different modes of expression support an increased interest in methods like project work (Kozma, 2003), which can be used for formative assessment. The increased use of digital portfolios in many countries (McFarlane, 2003) is an example of how formative assessment is gaining importance. Although the use of portfolio assessments is not new and has been used for some time without ICT (see e.g., special issue in *Assessment in Education*, 1998, on portfolios and records of achievement, Koretz, Broadfoot & Wolf, 1998), the use of digital tools seems to have developed this type of assessment further by bringing in some new qualitative dimensions such as possibilities for sending files electronically, hypertexts with links to other documents, and multimodality with written text, animations, simulations, moving images, and so forth. As a tool for formative assessment, and compared with paper-based portfolios, digital portfolios make it easier for teachers to keep track of documents, follow students' progress, and comment on students' assignments.. In addition, digital portfolios are used for summative assessment as documentation of the product students have developed and their progress. This offers greater choice and variety in the reporting and presenting of student learning (Woodward & Nanlohy, 2004). This research indicates a strengthening of collaboration (teamwork) and self-regulated learning skills. Related research deals with critical thinking skills. an area of student skill highlighted in curricula in many countries. What is needed in the application of ICT to assessment is to look for new ways of making student attainment visible in a valid and reliable way (Gipps & Stobart, 2003). (See also Thai school project, critical thinking skills, Rumpagaporn & Darmawan, 2007.)

In short, in the matter of measuring more traditional skills, development has been directed toward the delivery of large-scales tests on information handling and mapping levels of knowledge at different stages of schooling. Information literacy in this sense has become an important area of skill in itself, and even more so in relation to information sources on the Internet. ICT is seen as an important tool in making assessment more efficient as well as more effective in measuring desired skills in traditional ways.

### **The transformational strategy with ICT**

Although there are few instances of transformative e-assessment, the projects that do exist provide us with a compelling case for researching and investing in assessments of this type. There are exciting and effective examples of the use of ICT to transform assessment (and thence, learning). What is changing in the e-assessment field is usability. Where previously much of the preparatory work had to be done by third party or other technically expert staff, programs are increasingly providing end-users with the tools to implement their own e-assessment. New technologies have created an interest in what some describe as "assessing the inaccessible" (Nunes et al., 2003), such as metacognition, creativity, communication, learning to learn and lifelong learning skills (Anderson, 2008; Deakin, Crick et al., 2004). Below we review the research on assessing complex skills that have been difficult to assess or not assessed at all with traditional tests.

The *Review of advanced e-assessment techniques* project – commissioned by the Joint Information Systems Committee (JISC) in the UK and led by Martin Ripley - began by considering what constituted an advanced technique. "Advanced" refers to techniques that are used in isolated or restricted domains, and which have successfully applied technology to create an assessment tool. "Advanced" does not necessarily imply "newness". The project collated a long-list of over 100 "advanced" e-assessment projects. It was a surprise how few previously unknown advanced e-assessment projects came to light through the trawls for information. The community of experts



Four ICT skills were assessed:

1. Finding things out – obtaining information well matched to purpose by selecting appropriate sources; or, questioning the plausibility and value of information found.
2. Developing ideas and making things happen – using ICT to measure, record, respond to and control events.
3. Exchanging and sharing information – using ICT to share and exchange information, such as web publishing and video conferencing.
4. Reviewing, modifying and evaluating work as it progresses – reflecting critically on own and others' use of ICT.

The design included a simulated environment in which students complete tests; a desktop environment with software and tools for students; new ways of scoring student performances based on the ICT processes students used to solve problems rather than the products, and new ways of enabling access to tests for all students. In one case, an email ostensibly from the editor of a local news-site would request students to research local job vacancies and prepare a vacancies page for the website. To complete this task, students would need to run web-searches and email virtual companies to request more information about vacancies. The extent and quality of information available would vary, reflecting real-world web information. While completing the task, a student would receive further requests from the editor, perhaps changing deadlines or adding requirements. A student's work would be graded automatically.

The project provided proof-of-concept and identified the following major obstacles and challenges in developing a simulation-based assessment of 21st century skills.

- Developing a psychometric approach to measuring and scaling student responses. Since the assessment is designed to collect information about processes used by students, a method is needed to collect data and create summary descriptions/analyses of those processes.
- Aligning schools' technology infrastructure to support wide-scale, high-stakes, computer-based testing.
- Communicating effectively in introducing new approaches to testing to a world of experts, teachers, students, parents and politicians, all of whom have their own mental models and classical approaches for evaluating tests.

### **Box 1: Innovative UK assessment of ICT skills of 14-year-olds**

using e-assessment is small, and this continues to have implications for scaling e-assessment and for stimulating the growth of additional innovative approaches. A brief description of an advanced e-assessment developed in the UK is provided in Box 1.

One important aspect about the advances in e-assessment is that ICT brings new dimensions to what is being measured. Consider for example multimodality, or what Gunter Kress describes as multimodal literacy (2003). How might different skills like creativity, problem solving and critical thinking be expressed in different ways using different modes and modalities that ICT provides? The increased uses of visualization and simulation are examples of how ICT has made an impact on measurement of different skills, though so far the research has been inconclusive (Wegerif & Dawes, 2004).

Creativity especially is an area that has been growing in importance as a key 21st century thinking skill (Wegerif & Dawes, 2004, p.57). For example, Web 2.0 technology enables users to produce and share content in new ways: user-generated content creation and 'remixing' (Lessig, 2008) become creative practices that challenge the traditional relationships between teachers and students in providing information and content for learning, and the role of the 'school book' (Erstad, 2008). The use of new digital media in education has been linked to assessment of creative thinking as different from analytic thinking (Ridgway et al., 2004). Digital camera and different software tools make it easier for students to show their work and reflect on it. However, one of the problems with the discussions around creativity has been the often simplified and naïve notions and romantic conceptions of the creative individual (Banaji & Burn, 2007), without clear specifications of what this skill area might entail. Thus, it has proved to be difficult to assess students' creativity. In a systematic review of the impact of the use of ICT on students and teachers for the assessment of creative and critical thinking skills, Harlen & Deakin Crick, (2003) argue that the neglect of creative

and critical thinking in assessment methods is a cause for concern, given the importance of these skills in the preparation for life in a rapidly changing society and for lifelong learning. Their review documents a lack of substantial research on these issues and argues for more strategic research.

A second area of great interest concerns the way digital tools can support collaboration in problem solving, creative practices, and communication. There are many examples of how computer-based learning environments for collaboration can work to stimulate student learning and the process of inquiry (Wasson et al., 2003; Laurillard, 2009). Collaborative problem-solving skills are considered necessary for success in today's world of work and school. Online collaborative problem-solving tasks offer new measurement opportunities when information on what individuals and teams are doing is synthesized along cognitive dimension. Students can send documents and files to each other and in this way work on tasks together. This raises issues both on interface design features that can support online measurement and how to evaluate collaborative problem-solving processes in an online context (O'Neil et al., 2003). There are also examples of web-based peer assessment strategies (Lee et al., 2006). Peer assessment has been defined by some as an innovative assessment method, since students themselves are put in the position of evaluators as well as learners (Lin et al., 2001). It has been used with success in different fields such as writing, business, science, engineering, and medicine.

A third area of research with great implications for how ICT challenges assessment concerns higher-order thinking skills. Ridgway & McCusker (2003) show how computers can make a unique contribution to assessment in the sense that they can present new sorts of tasks, whereby dynamic displays show changes in several variables over time. The authors cite examples from the World Class Arena ([www.worldclassarena.org](http://www.worldclassarena.org)) to demonstrate how these tasks and tools support complex problem solving for different age groups. They show how computers can facilitate the creation of micro-worlds for students to explore in order to discover hidden rules or relationships, like virtual laboratories for doing experiments or games to explore problem-solving strategies. Computers allow students to work with complex data sets of a sort that would be very difficult to work with on paper. Tools like computer-based simulations can in this way give a more nuanced understanding of what students know and can do than traditional testing methods (Bennett et al., 2003). Findings such as those reported by Ridgway & McCusker (2003) are positive in the way students relate to computer-based tasks and the increased performances they exhibit. However, the authors also find that students have problems in adjusting their strategies and skills since the assessment results shows that they are still tuned into the old test situation with correct answers rather than explanations and reasoning skills.

An interesting new area associated with what has been presented above is the knowledge building perspective developed by Scardamalia and Bereiter (2006). (See also White Paper 4.) In developing the technological platform *Knowledge forum*, Scardamalia and Bereiter have been able to measure students learning processes that traditionally have been difficult to assess. This platform gives the students the possibility of collective reasoning and problem solving building on each other's notes, often as collaboration between schools in different sites and countries. Some key themes in the research on these skills and their online measurement tools are:

- Knowledge advancement as a community rather than individual achievement.
- Knowledge advancement as idea improvement rather than as progress toward true and warrant belief.
- Knowledge of in contrast to knowledge about.
- Discourse as collaborative problem solving rather than as argumentation.
- Constructive use of authoritative information.
- Understanding as emergent.

Similar points have been made by Mercer and Wegerif and colleagues in the UK (e.g. Mercer & Littleton, 2007) in their research on 'thinking together' (<http://thinkingtogether.educ.cam.ac.uk/>) and how we might build language for thinking, what they term "exploratory talk". Computers and software have been developed for this purpose together with other resources. Wegerif and Dawes

(2004: 59) have summarized the “thinking together” approach in four points each of which assumes the crucial importance of teachers:

- The class undertakes explicit teaching and learning of talk skills that promote thinking.
- Computers are used both to scaffold children’s use of these skills and to bridge them in curriculum areas.
- Introductions and closing plenaries are used to stress aims for talk and for thinking as well as to review progress.
- Teacher intervention in group work is used to model exploratory talk.

The above examples have shown how ICT represents the transformative strategy in developing assessments, especially formative assessment, and how the complexity of these tools can be used to assess skills that are difficult to assess by paper and pencil. As McFarlane (2001) notes, “It seems that use of ICT can impact favourably on a range of attributes considered desirable in an effective learner: problem-solving capability; critical thinking skill; information-handling ability” (p. 230). Such skills can be said to be more relevant to the needs in the information society and the emphasis on lifelong learning than those which traditional tests and paper-based assessments tend to measure.

## **Arriving at a model 21st century skills framework and assessment**

In this section we provide a framework that could be used as a model for developing large-scale assessments of 21st century skills. To arrive at this model framework, we compared a number of available curriculum and assessment frameworks for 21st century skills and skills that have been developed around the world. We analyzed these frameworks to determine not only the extent to which they differ, but also the extent to which these frameworks provide descriptions of 21st century learning outcomes in measureable form. Based on our analysis, we identified 10 important skills (or skills) that in our opinion typify the skills necessary for the 21st century. For each of the 10 skills, we have analyzed the extent to which the identified frameworks provide measureable descriptions of the skill, considering the **K**nowledge, **S**kills, and **A**ttitudes, **V**alues and **E**thics aspects of each skill. This framework is referred to as the **KSAVE** framework, and is described in more detail below.

## **Existing 21st century skills frameworks**

A number of organizations around the world have independently developed frameworks for 21st century skills and skills. For the purposes of our analysis we considered the frameworks listed in the chart appearing on the next page. To explore the number and range of modern 21st century curricula that are currently in place, wider searches were carried out for national education systems that build aspects of the 10 KSAVE skills into their national curricula. Searches were made for “national” curricula, references to “21st century learning” and references to “skills” and “competency based” standards. A relatively small number of nations define a national curriculum in detail while a larger number have national aims or goals for their education system. A growing number of countries are undertaking significant reviews of their national curricula; a very few are undertaking the task of developing their first national curriculum. ‘21st century learning needs’ are frequently included within these new and revised curriculum documents. The sources are listed in Table 1.

With very few exceptions, references to 21st century knowledge, skills or the individual attitudes and attributes of learners are contained within overarching statements of goals or educational aims. These are generally brief statements, but are supported by justifications for change. For example, there are references to the need to educate for new industry, commerce, technology and economic structures; the need for new social interaction and communication skills; the need for imagination, creativity and initiative; the need to learn and continue to learn throughout employment; the need to maintain national and cultural values; and the need to operate in an increasingly international and global environment. Few of the frameworks and curricula of national systems we have examined provide detailed descriptions, or clearly elaborated curriculum standards. Similarly few include

**Table 1: Sources of documents on 21st century skills**

Country/ Region	Document(s)
European Union	<i>Key Competencies for Lifelong Learning – A European Reference Framework, November 2004</i> Recommendation of the European Parliament and of the Council of 18 December 2006 on key competences for lifelong learning. Implementation of “Education and Training 2010” Work Programme. <a href="http://ec.europa.eu/dgs/education_culture/.../ll-learning/keycomp_en.pdf">http://ec.europa.eu/dgs/education_culture/.../ll-learning/keycomp_en.pdf</a>
OECD	New Millennium Learners Project: Challenging our Views on ICT and Learning <a href="http://www.oecd.org/document/10/0,3343,en_2649_35845581_38358154_1_1_1_1,00.html">www.oecd.org/document/10/0,3343,en_2649_35845581_38358154_1_1_1_1,00.html</a>
USA (Partnership for 21st Century Skills)	P21 Framework Definitions P21 Framework Flyer <a href="http://www.21stcenturyskills.org/documents/p21_framework_definitions_052909.pdf">www.21stcenturyskills.org/documents/p21_framework_definitions_052909.pdf</a>
Japan	Center for Research on Educational Testing (CRET) Source: ??
Australia	<i>Melbourne declaration on educational goals for young Australians</i> <a href="http://www.mceecdya.edu.au/verve/_resources/National_Declaration_on_the_Educational_Goals_for_Young_Australians.pdf">www.mceecdya.edu.au/verve/_resources/National_Declaration_on_the_Educational_Goals_for_Young_Australians.pdf</a>
Scotland	A Curriculum for Excellence – the Four Capabilities <a href="http://www.ltscotland.org.uk/curriculumforexcellence/index.asp">www.ltscotland.org.uk/curriculumforexcellence/index.asp</a>
England	<i>The learning journey</i>
England	<i>Personal learning &amp; thinking skills – The national curriculum for England</i> <a href="http://curriculum.qca.org.uk/uploads/Personal,%20learning%20and%20thinking%20skills%20leaflet_tcm8-12831.pdf">http://curriculum.qca.org.uk/uploads/Personal,%20learning%20and%20thinking%20skills%20leaflet_tcm8-12831.pdf</a>
Northern Ireland	<i>Assessing the cross curricular skills</i> <a href="http://www.nicurriculum.org.uk/key_stage_3/assessment_and_reporting/cross_curricular_skills.asp">www.nicurriculum.org.uk/key_stage_3/assessment_and_reporting/cross_curricular_skills.asp</a>
ISTE	<i>National educational technology standards for students, second edition, global learning in the digital age</i> Source: ??
USA. National Academies, science for the 21st century	<i>Exploring the intersection of science education and the development of 21<sup>st</sup> century skills</i> Source: ??
USA, Department of Labor	Competency models: <i>A review of the literature</i> <i>The role of the Employment and Training Administration (ETA), Michelle R. Ennis</i>

descriptions of what the curriculum experienced by learners will actually look like if the broader aims of your framework are to be realized.

All the curricula reviewed maintain a subject structure and it is this structure that forms the basis for curriculum design. The naming and grouping of learning under subject titles may differ slightly between countries but the general principles of learning a core curriculum (a home language, mathematics, and science) are common. In many national curricula the skills associated with ICT have been raised in status to this core while history, particularly the national history, and indigenous culture, often including religion, form a secondary layer. Other subjects may be described individually or combined, for example as the “Arts” or “Humanities”. Thus to date, the teaching of 21st century skills has been embedded in the subjects that make up the school curriculum. It is not even clear whether such skills as critical thinking or creativity have features in common in related subjects such as mathematics and science, let alone across the STEM fields and the arts and humanities. For other skills, however, such as information and ICT literacy, the argument has been

made more frequently that these are transferrable. These questions of skill generalizability and transferability remain as deep research challenges.

Where the aims and goals of 21st century learning are described in the frameworks we examined, these are generally specified as being taught through, within and across the subjects without the detail of how this is to be achieved or what the responsibilities of each subject might be in achieving them. Without this depth of detail these national statements of 21st century aims and goals are unlikely to be reflected in the actual learning experience of students or in the assessments that are administered. Without highly valued assessments of these 21st century aims or goals requiring their teaching, it is difficult to see when or how education systems will change significantly for the majority of learners.

## **The KSAVE model**

To structure the analysis of 21st century skills frameworks, an overall conceptual diagram was created. This diagram defines 10 skills grouped into 4 categories, as follows:

### **Ways of Thinking**

5. Creativity and innovation
6. Critical thinking, problem solving, decision making
7. Learning to learn, metacognition

### **Ways of Working**

8. Communication
9. Collaboration (teamwork)

### **Tools for Working**

10. Information literacy (includes research on sources, evidence, biases, etc.)
11. ICT literacy

### **Living in the World**

12. Citizenship – local and global
13. Life and career
14. Personal & social responsibility – including cultural awareness and competence

Although there are significant differences in the ways in which these skills are described and clustered from one framework to another, we consider that the above list of ten is sufficiently broad and comprehensive to accommodate all approaches. At an early stage, we found that frameworks for 21st century skills differ considerably in terms of the nature of their content. Some seek to define student behaviors. For example, an aspect of creativity might include “openness and responsiveness to new ideas”. Other frameworks refer extensively to skills: for example, an aspect of creativity might refer to the ability to “develop innovative and creative ideas”. A third category used in some frameworks refers to specific knowledge: for example, an aspect of creativity might be “knowledge of a wide range of idea creation techniques.” Some frameworks cover two or more of these categories; few comprehensively cover all three. To accommodate and reflect these differences in approach, we have designed three categories within the KSAVE model. Keep in mind that the model does not resolve the issue of subject-embedded knowledge, skills and attitudes versus their generalizability across domains.

### **Knowledge**

This category includes all references to specific knowledge or understanding requirements for each of the 10 skills.

### **Skills**

This category includes the abilities, skills and processes that curriculum frameworks are designed to develop in students and which are a focus for learning.

**Table 2: Ways of thinking - creativity and innovation**

Knowledge	Skills	Attitudes/Values/Ethics
<p>Think and work creatively and with others</p> <ul style="list-style-type: none"> <li>• Know a wide range of idea creation techniques (such as brainstorming)</li> <li>• Be aware of invention, creativity and innovation from the past within and across national boundaries and cultures</li> <li>• Know the real world limits to adopting new ideas and how to present them in more acceptable forms</li> <li>• Know how to recognize failures and differentiate between terminal failure and difficulties to overcome</li> </ul> <p>Implement innovations</p> <ul style="list-style-type: none"> <li>• Be aware of and understand where and how innovation will impact and the field in which the innovation will occur</li> <li>• Be aware of the historical and cultural barriers to innovation and creativity</li> </ul>	<p>Think creatively</p> <ul style="list-style-type: none"> <li>• Create new and worthwhile ideas (both incremental and radical concepts)</li> <li>• Be able to elaborate, refine, analyze and evaluate one's own ideas in order to improve and maximize creative efforts</li> </ul> <p>Work creatively with others</p> <ul style="list-style-type: none"> <li>• Develop, implement and communicate new ideas to others effectively</li> <li>• Be sensitive to the historical and cultural barriers to innovation and creativity</li> </ul> <p>Implement Innovations</p> <ul style="list-style-type: none"> <li>• Develop innovative and creative ideas into forms that have impact and be adopted</li> </ul>	<p>Think creatively</p> <ul style="list-style-type: none"> <li>• Be open to new and worthwhile ideas (both incremental and radical concepts)</li> </ul> <p>Work creatively with others</p> <ul style="list-style-type: none"> <li>• Be open and responsive to new and diverse perspectives; incorporate group input and feedback into the work</li> <li>• View failure as an opportunity to learn; understand that creativity and innovation is a long-term, cyclical process of small successes and frequent mistakes</li> </ul> <p>Implement Innovations</p> <ul style="list-style-type: none"> <li>• Show persistence in presenting and promoting new ideas</li> </ul>

### Attitudes, Values, Ethics

This category refers to the behaviors and aptitudes that students exhibit in relation to each of the 10 skills.

The methodology used to complete the analysis of 21st century skills frameworks was to populate the KSAVE grid with indexes taken from each framework; retaining original wording as far as was sensible. Decisions were made to refine or amalgamate wording taken from frameworks where the intention appeared similar. Decisions were also made on whether to allocate indexes to knowledge, skills, or attitudes/values/ethics. For some of the indexes, the decision whether to allocate them to the skills category or to the attitudes/values/ethics category appeared to be marginal.

In the following pages we present each group of skills and discuss some of the thinking behind the grouping. In addition, we provide examples of how the skills might be measured in an effort to open our eyes to what is possible. These example assessments really only scratch the surface of what is needed to measure 21st century skills.

## Ways of thinking

Together the three categories of skills under *Ways of thinking* represent a push forward in the conceptualization of thinking. These skills emphasize the upper end of thinking skills, and subsume more straightforward skills such as recall, and drawing inferences. A major characteristic of these skills is that they require greater focus and reflection.

### Creativity and Innovation

Operational definitions of creativity and innovation are provided in Table 2. While creativity and innovation can logically be grouped together, they originate in two different traditional schools of

thought. Creativity is most often the concern of cognitive psychologists. Innovation, on the other hand, is more closely related to economics where the goal is to improve, advance, and implement new products and ideas. Measuring both can be quite challenging. Not only do the tasks require an interactive environment, but also they frequently cannot be done in the short period of time allocated to a large-scale assessment, nor are there good benchmarks against which the respondent output can be evaluated.

Creativity is often described as a thinking skill or at least as an important aspect of thinking that can and should be fostered (Wegerif & Dawes, 2004, p.57). In a review of the interconnection between technology, learning and creativity, Loveless (2002) shows how technology allows children, quickly and easily, to produce high quality finished products in a range of media that provide opportunities for creativity. Loveless argues that to foster creativity in the classroom, teachers need to create a social atmosphere in which children feel secure enough to play with ideas and to take risks.

Although as noted above, it has proved to be difficult to assess creativity, the use of new digital media has been linked to assessment of creative thinking as different from analytic thinking (Ridgway et al., 2004). Digital camera and different software tools make it easier for students to show their work and reflect on it. A number of subjects in the school curriculum ask students to make various kinds of products. (Sefton-Green & Sinker, 2000). These might include paintings in art, creative writing in English, performance in drama, recording in music, videos in media studies, and multimedia “digital creations” in different subjects. There are so far not many examples of how ICT might influence assessment of such student products (Sefton-Green & Sinker, 2000).

### eSCAPE

The eSCAPE project does not test creativity and innovation but it does test some aspects of this domain. Specifically it offers a glimpse of how we might test the ability to develop innovative and creative ideas into forms that have impact and be adopted as well as showing persistence in presenting and promoting new ideas

For many years, England's school examinations for 16 year-old students have included an optional assessment in Design and Technology. Traditionally this examination includes a requirement for students to complete a design project of over 100 hours duration, and for the student to submit a written report on the project. The report is graded.

In 2001, the Qualifications and Curriculum Authority commissioned the Technology Education Research Unit (TERU) at Goldsmiths College in London to undertake to develop a technology-led replacement to this traditional paper-based assessment. The result is an assessment completed in 6 hours, in a design workshop, with students working in groups of 3 or 4. During the course of the 6 hours, students are given a number of staged assessment instructions and information via a personal, portable device. The handheld device also acts as the tool to capture assessment evidence – via video, camera, voice, sketchpad and keyboard. During the 6 hours, each student's design prototype develops, with the handheld device providing a record of progress, interactions and self-reflections.

At the end of the assessment, the assessment evidence is collated into a short multi-media portfolio. Human raters, who score each student's responses, view this. eSCAPE directors turned to the work of Thurstone to develop a graded-pairs scoring engine to provide a holistic judgment on the students' work,. This engine supports human raters in making a number of paired judgments about students' work. The result is an assessment that exhibits rates of reliability equal to, or slightly in excess of, the levels of reliability achieved on multiple-choice tests.

### *Critical thinking, problem solving and decision making*

Operational definitions of critical thinking and problem solving are provided in Table 3. Critical thinking and problem solving have become an increasingly important feature of the curriculum in

**Table 3: Ways of thinking - critical thinking, problem solving, decision making**

Knowledge	Skills	Attitudes/Values/Ethics
Reason effectively, use systems thinking and evaluate evidence <ul style="list-style-type: none"> <li>• <i>Understand systems and strategies for tackling unfamiliar problems</i></li> <li>• <i>Understand the importance of evidence in belief formation</i></li> </ul> Reevaluate beliefs when presented with conflicting evidence Solve problems <ul style="list-style-type: none"> <li>• <i>Identify gaps in knowledge</i></li> <li>• <i>Ask significant questions that clarify various points of view and lead to better solutions</i></li> </ul> Articulation <ul style="list-style-type: none"> <li>• <i>Clearly articulate the results of one's inquiry</i></li> </ul>	Reason effectively <ul style="list-style-type: none"> <li>• <i>Use various types of reasoning (inductive, deductive, etc.) as appropriate to the situation</i></li> </ul> Use systems thinking <ul style="list-style-type: none"> <li>• <i>Analyze how parts of a whole interact with each other to produce overall outcomes in complex systems. Examine ideas, identify and analyse arguments.</i></li> <li>• <i>Synthesize and make connections between information and arguments</i></li> <li>• <i>Interpret information and draw conclusions based on the best analysis. Categorise, decode and clarify information.</i></li> <li>• <i>Effectively analyze and evaluate evidence, arguments, claims and beliefs</i></li> <li>• <i>Analyze and evaluate major alternative points of view</i></li> <li>• <i>Evaluate. Assess claims and arguments.</i></li> <li>• <i>Infer. Query evidence, conjecture alternatives and draw conclusions.</i></li> <li>• <i>Explain. Stating results, justifying procedures and presenting arguments.</i></li> <li>• <i>Self-regulate, self-examine and self correct.</i></li> </ul>	Make reasoned judgments and decisions <ul style="list-style-type: none"> <li>• <i>Consider and evaluate major alternative points of view</i></li> <li>• <i>Reflect critically on learning experiences and processes</i></li> <li>• <i>Incorporate these reflections into the decision-making process.</i></li> </ul> Solve problems <ul style="list-style-type: none"> <li>• <i>Be open to non-familiar, unconventional and innovative solutions to problems and to ways to solve problems</i></li> <li>• <i>Ask meaningful questions that clarify various points of view and lead to better solutions</i></li> </ul> Attitudinal disposition <ul style="list-style-type: none"> <li>• <i>Trustful of reason</i></li> <li>• <i>Inquisitive and concerned to be well informed</i></li> <li>• <i>Open and fair minded</i></li> <li>• <i>Flexible and honest</i></li> <li>• <i>Inquisitiveness and concern to be well informed</i></li> <li>• <i>Alert to opportunities to Use ICT</i></li> <li>• <i>Trustful of and confident in reason</i></li> <li>• <i>Open and fair minded, flexible in considering alternative opinions</i></li> <li>• <i>Honest assessment of one's own biases</i></li> <li>• <i>Willingness to reconsider or revise one's views where warranted.</i></li> </ul>

many parts of the world. In the UK, there are popular high school qualifications in critical thinking. In the USA, the American Philosophical Association has published the Delphi report on critical thinking (Facione, 1990). This report identified six cognitive thinking skills: interpretation, analysis, evaluation, inference, explanation and self-regulation. This framework was further elaborated to include attitudinal and values based criteria: students should be inquisitive, well informed, open-minded, fair, flexible and honest. Research subsequent to the Delphi Report has shown that being "trustful of reason" (one of the Delphi Report's key findings) plays a vital role in what it means to think critically.

In contrast to creativity and innovation, critical thinking, problem solving, and decision-making have been part of large-scale assessments for sometime. Critical thinking frequently appears as part of reading, mathematics and science assessments, with such assessments as the U.S. National Assessment of Educational Progress and the OECD Program for International Student Achievement (PISA).

Problem solving has been a focused area of research for decades yielding a number of definitions and frameworks. In addition, problem solving has appeared in various forms in a number of large-



scale international assessments such as PISA and the Adult Literacy and Lifelong Learning Skills (ALL). These assessments specifically include items that are designed to measure how well students can evaluate evidence, arguments, claims and warrants; synthesize and make connections between information and arguments; and analyze and evaluate alternative points of view. ALL 2003 focused on problem solving tasks that were project oriented and most closely resembled analytic reasoning. Problem solving in mathematics and science has been part of the PISA assessment since its inception in 2000. In PISA 2003 a problem solving scale that included 3 kinds of problems – decision-making, system analysis and design, and troubleshooting was developed. For 2012 PISA will move beyond the 2003 scale by include dynamic items that may be linked to PIAAC 2011 where problem solving in a technology rich environment.

The following examples illustrate the direction of assessments for the 21st century. The first, Primum, from the USA, illustrates authentic open-ended tasks that can be machine scored. The second example, World Class Tests, illustrates highly innovative problem solving in mathematics, science, and design and technology that are, by design, non-familiar to the student (much of our current testing is routine and predictable), interesting and motivating, psychologically challenging, and focused on a specific dimension of problem solving (such as optimization, or visualization) in a math/science/design context. These tasks offer the hope that it is possible to design lively 5-10 minute long, interactive and complex problems for students to solve in the context of an on-screen test. The third example, the Virtual Performance Assessment (VPA) project, also from the USA, addresses the feasibility of using immersive technologies to deliver virtual performance assessments that measure science inquiry knowledge and skills, as defined by the US National Science Education Standards (NRC, 1996).

#### Primum

Some advocates of e-assessment point to the potential of computers to support simulation and scenario-based assessment. There are few examples of this category of e-assessment being developed successfully, especially not in high stakes testing contexts. Primum, which assesses decision-making in a very specific context, is an exception. It provides an assessment of trainee medical practitioners' ability to make medical diagnoses when presented with a fictitious patient exhibiting a number of symptoms. This automated assessment has been designed to provide an authentic and reliable assessment at a price that compares favorably with the alternative – human scored evaluation at patients' bedsides.

#### World Class Tests

In 2000 England's Department for Education commissioned the development of new computer-based tests of problem solving, in the domains of mathematics, science, design and technology. These tests are intended for worldwide application; they were designed to make creative use of computer technology. Also, they are intended to set new benchmarks in the design of assessments of students' thinking and ability to apply a range of techniques to solve novel and unexpected problems. These tests have become known as World Class Tests and have been adapted for children aged 8-14. These tests are now sold commercially under license in the Far East.

#### The VPA Project

The Virtual Performance Assessment project utilizes innovations in technology and assessment to address the problem of measuring a student's ability to perform scientific inquiry to solve a problem. The project is developing assessments for use in school settings as a standardized component of an accountability program. The goal is to develop three assessments in the context of life science that appear different on the surface, but are all measuring the same inquiry process skills. Each assessment will take place in a different type of ecosystem, and students will investigate authentic ecological problems as they engage in the inquiry process.

**Table 4: Ways of thinking - learning to learn, metacognition**

Knowledge	Skills	Attitudes/Values/Ethics
<ul style="list-style-type: none"> <li>• <i>Knowledge and understanding of one's preferred learning methods, the strengths and weaknesses of one's skills and qualifications.</i></li> <li>• <i>Knowledge of available education and training opportunities and how different decisions during the course of education and training lead to different careers.</i></li> </ul>	<ul style="list-style-type: none"> <li>• <i>Effective self-management of learning and careers in general: ability to dedicate time to learning, autonomy, discipline, perseverance and information management in the learning process.</i></li> <li>• <i>Ability to concentrate for extended as well as short periods of time.</i></li> <li>• <i>Ability to reflect critically on the object and purpose of learning.</i></li> <li>• <i>Ability to communicate as part of the learning process by using appropriate means (intonation, gesture, mimicry, etc.) to support oral communication as well as by understanding and producing various multimedia messages (written or spoken language, sound, music etc.).</i></li> </ul>	<ul style="list-style-type: none"> <li>• <i>A self-concept that supports a willingness to change and further develop skills as well as self-motivation and confidence in one's capability to succeed.</i></li> <li>• <i>Positive appreciation of learning as a life-enriching activity and a sense of initiative to learn.</i></li> <li>• <i>Adaptability and flexibility.</i></li> <li>• <i>Identify personal biases.</i></li> </ul>

### *Learning to learn and metacognition*

Operational definitions of learning to learn and metacognition are provided in Table 4. Learning to learn and metacognition have most frequently been measured by think aloud protocols that have been administered in one-on-one situations. Clearly this methodology is not amenable to large-scale assessments. However, technology might be used to support and assess 'learning to learn', which includes self-assessment and self-regulated learning. One interesting example of this is the eVIVA-project developed at Ultralab in the United Kingdom.

### eVIVA

The intention of eViva was to create a more flexible way of assessment, taking advantage of the possibilities new technologies such as a mobile phone and web-based formative assessment tools offer. By using such tools Ultralab promoted self- and peer-assessment as well as dialogue between teachers and students.

In this project the students had access to the eVIVA website where they could set up an individual profile of system preferences and recording an introductory sound file, on their mobile or land phone. After this students' could then carry out a simple self-assessment activity by selecting a series of simple "I Can" statements designed to start them thinking about what they are able to do in ICT. The website consisted of a question bank from which the pupils were asked to select 4 or 5 questions for their telephone viva or assessment carried out toward the end of their course, but at a time of their own choice. Students were guided in their choice by the system and their teacher. They had their own e-portfolio web-space in which they were asked to record significant *milestone* moments of learning, and to upload supporting files as evidence. Each milestone was then annotated or described by the pupil to explain what they had learned or why they were proud of a particular piece of work. Once milestones had been published, teachers and pupils could use the annotation and the messaging features to engage in dialogue with each other about the learning. Students were encouraged to add comments to their own and each other's work and the annotations could be sent via phone using SMS or voice messages. When ready, students would dial into eVIVA, either by mobile or land phone, and record their answers to their selected

questions. This gave students the opportunity to explain what they had done and reflect further on their work. Their answers were recorded and sent to the website as separate sound files. The teacher made a holistic assessment of the pupil's ICT capabilities based on the milestones and work submitted in the e-portfolio, student reflections or annotations, the recorded eVIVA answers and any written answers attached to the questions, and classroom observations (see Walton, 2005).

### Cascade

Cascade, which is under development at the University of Luxembourg and the Center for Public Research Henri Tudor, is an innovative item type that is more amenable to large-scale assessments with limited testing time.

The Cascade test items are designed so that respondents answer a set of questions and then are asked to rate how certain they are about the correctness of their response on each item. Then the respondent is given an opportunity to access multimedia information to verify the correctness of the response. At that point the respondent once again answers the same set of questions and again rates his/her certainty. Scoring is based on the comparison of the first and second set of responses and using trace information the paths that he/she took in acquiring additional information.

### **Ways of working**

In business we are witnessing a rapid shift in the way people work. Outsourcing services across national and continental borders is just one example. Another is having team members telecommute while working on the same project. For instance a small software consulting team has members located on three continents. They work on developing prototypes using teleconferences, text messaging, with the occasional "sprint" sessions where they gather in a single location and work 24 hours a day to develop the product. Similarly, in the large-scale international assessments such as PISA, TIMSS (Trends in Mathematics and Science Study), and PIAAC, teams of researchers and developers across continents and at multiple locations work together to develop the assessments. To support these examples of multiple moves toward globalization, communication and collaboration skills must be more finely honed. Communication must be rapid, concise and cognizant of cultural differences.

### *Communication*

Operational definitions of communication are provided in Table 5. Communication has been a mainstay of assessments in the form of reading, writing and alternative representations such as graphing in mathematics and science, listening and speaking, the assessments have not taken into account the full range of possibilities. At the most minimal, PowerPoint presentations are now ubiquitous. These frequently include graphic displays that in conjunction with language can more succinctly deliver the message. Video presentations also require the combination of communication forms in ways that have never before been within the realm of almost everyone's capability. To date newer modes of communication have rarely been represented in large-scale assessments. However, in light of the developments described below it is essential that we take these changes into account.

Consider the use of text messaging. The first commercial text message was sent in December of 1992. Today the number of text messages sent and received everyday exceeds the total population of the planet. Facebook that started as a communication vehicle for college students reached a market audience of 50 million people within just 2 years. Currently Facebook has more than 200 million active users and more than 100 million users log on at least once each day. It has now moved into business applications, with business and interest groups having Facebook pages. According to Business Week (Aug 2007) there has been a 113% rise since 2006 in the number of Facebook visitors over the age of 35. It is also increasingly more common to use Facebook as the venue for organizing and conducting conferences.

**Table 5: Ways of working - communication**

Knowledge	Skills	Attitudes/Values/Ethics
<p>Competency in language in mother tongue.</p> <ul style="list-style-type: none"> <li>• <i>Sound knowledge of basic vocabulary, functional grammar and style, functions of language.</i></li> <li>• <i>Awareness of various types of verbal interaction (conversations, interviews, debates, etc.) and the main features of different styles and registers in spoken language.</i></li> <li>• <i>Understanding the main features of written language (formal, informal, scientific, journalistic, colloquial, etc.).</i></li> </ul> <p>Competency in additional language/s.</p> <ul style="list-style-type: none"> <li>• <i>Sound knowledge of basic vocabulary, functional grammar and style, functions of language.</i></li> <li>• <i>Understanding the paralinguistic features of communication (voice-quality features, facial expressions, postural and gesture systems).</i></li> <li>• <i>Awareness of societal conventions and cultural aspects and the variability of language in different geographical, social and communication environments.</i></li> </ul>	<p>Competency in language in mother tongue and additional language/s.</p> <ul style="list-style-type: none"> <li>• <i>Ability to communicate, in written or oral form, and understand, or make others understand, various messages in a variety of situations and for different purposes.</i></li> <li>• <i>Communication includes the ability to listen to and understand various spoken messages in a variety of communicative situations and to speak concisely and clearly.</i></li> <li>• <i>Ability to read and understand different texts, adopting strategies appropriate to various reading purposes (reading for information, for study or for pleasure) and to various text types.</i></li> <li>• <i>Ability to write different types of texts for various purposes. To monitor the writing process (from drafting to proof-reading).</i></li> <li>• <i>Ability to formulate one's arguments, in speaking or writing, in a convincing manner and take full account of other viewpoints, whether expressed in written or oral form.</i></li> <li>• <i>Skills needed to use aids (such as notes, schemes, maps) to produce, present or understand complex texts in written or oral form (speeches, conversations, instructions, interviews, debates).</i></li> </ul>	<p>Competency in language in mother tongue.</p> <ul style="list-style-type: none"> <li>• <i>Development of a positive attitude to the mother tongue, recognizing it as a potential source of personal and cultural enrichment.</i></li> <li>• <i>Disposition to approach the opinions and arguments of others with an open mind and engage in constructive and critical dialogue.</i></li> <li>• <i>Confidence when speaking in public.</i></li> <li>• <i>Willingness to strive for aesthetic quality in expression beyond the technical correctness of a word/phrase.</i></li> <li>• <i>Development of a love of literature.</i></li> <li>• <i>Development of a positive attitude to intercultural communication.</i></li> </ul> <p>Competency in additional language/s.</p> <p><i>Sensitivity to cultural differences and resistance to stereotyping.</i></p>

Why are these communication innovations important? Beginning with text messaging we need to consider the shift in grammar, syntax and spelling that pervades these communications. If we consider the proliferation of videos on You Tube, it is important to see how effective different presentation forms of the same information can be. Similarly, Facebook presents even more challenges as it merges registers -- here professional and personal communications can exist side-by-side.

One prominent example of incorporating new technologies into measures of communication was developed for PISA 2009. PISA's Electronic Reading Assessment simulated reading in a web environment. In many ways this step forward represents not only migration to newer innovative assessment items, but it also represents a first step in transforming assessments to more authentic and up-to-date tasks.

**Table 6: Ways of working - collaboration, teamwork**

Knowledge	Skills	Attitudes/Values/Ethics
Interact effectively with others <ul style="list-style-type: none"> <li>• <i>Know when it is appropriate to listen and when to speak</i></li> </ul> Work effectively in diverse teams <ul style="list-style-type: none"> <li>• <i>Know and recognize the individual roles of a successful team and know own strengths and weaknesses recognizing and accepting them in others</i></li> </ul> Manage projects <ul style="list-style-type: none"> <li>• <i>Know how to plan, set and meet goals and to monitor and re-plan in the light of unforeseen developments</i></li> </ul>	Interact effectively with others <ul style="list-style-type: none"> <li>• <i>Speak with clarity and awareness of audience and purpose. Listen with care, patience and honesty</i></li> <li>• <i>Conduct themselves in a respectable, professional manner</i></li> </ul> Work effectively in diverse teams <ul style="list-style-type: none"> <li>• <i>Leverage social and cultural differences to create new ideas and increase both innovation and quality of work</i></li> </ul> Manage projects <ul style="list-style-type: none"> <li>• <i>Prioritize, plan and manage work to achieve the intended group result</i></li> </ul> Guide and lead others <ul style="list-style-type: none"> <li>• <i>Use interpersonal and problem-solving skills to influence and guide others toward a goal</i></li> <li>• <i>Leverage strengths of others to accomplish a common goal</i></li> <li>• <i>Inspire others to reach their very best via example and selflessness</i></li> <li>• <i>Demonstrate integrity and ethical behavior in using influence and power</i></li> </ul>	Interact effectively with others <ul style="list-style-type: none"> <li>• <i>Know when it is appropriate to listen and when to speak</i></li> <li>• <i>Conduct themselves in a respectable, professional manner</i></li> </ul> Work effectively in diverse teams <ul style="list-style-type: none"> <li>• <i>Show respect for cultural differences and be prepared to work effectively with people from a range of social and cultural backgrounds</i></li> <li>• <i>Respond open-mindedly to different ideas and values</i></li> </ul> Manage projects <ul style="list-style-type: none"> <li>• <i>Persevere to achieve goals, even in the face of obstacles and competing pressures</i></li> </ul> Be responsible to others <ul style="list-style-type: none"> <li>• <i>Act responsibly with the interests of the larger community in mind</i></li> </ul>

### *Collaboration and teamwork*

Operational definitions of communication are provided in Table 6. Collaboration presents a different set of challenges for large-scale assessments. At the most basic, school level assessments are focused on getting measures of individual performance. Consequently when faced with a collaborative task the most important question is how to assign credit to each member of the group, as well as how to account for differences across groups that may bias a given student's performance. This becomes an even larger issue within international assessments where cultural boundaries are crossed. For example ALL researched the potential for measuring teamwork. While the designers could generate teamwork tasks, at that time accounting for cultural differences became the insurmountable obstacle.

Several important research initiatives have worked on getting measures of individual performance that address key components of collaboration and measurement (Laurillard, 2009). For example Cakir, Zemel and Stahl (2009) have shown how group participants, in order to collaborate effectively in group discourse on a topic like mathematical patterns, must organize their activities in ways that share the significance of their utterances, inscriptions, and behaviors. Their analysis reveals methods by which the group co-constructs meaningful inscriptions in the interaction spaces of the collaborative environment. The integration of graphical, narrative, and symbolic semiotic modalities facilitates joint problem solving. It allows group members to invoke and operate with multiple realizations of their mathematical artifacts, a characteristic of deep learning of mathematics. Other research shows how engaging in reflective activities in interaction, such as explaining, justifying and evaluating problem solutions collaboratively can potentially be productive for learning (Baker & Lund, 1997). Several studies have also shown how taking part in collaborative inquiry

toward advancing a shared knowledge object can serve as a means to facilitate the development of metaskills.

Two further lines of research are pertinent to including collaborative work in large-scale assessments. The first line of research begins with the idea of a simulation where one respondent interacts with pre-programmed virtual partners. The drawback here is the current lack of theoretic understandings of how collaborators would interact in this environment. The second line of research is best exemplified by group tasks where evidence of interaction patterns and self-reflections are captured. Research into how to rate these interactions would lead to a rubric that might either be criterion-referenced or be normed according to country, nationality, SES, or other differentiating group characteristics. In conjunction with the product scores it would be possible to generate a collaboration scale on the basis of such research.

It has been observed that as employers we most often base our staff recruitment decisions on formal, school and college-based qualifications, using these as a measure of an applicant's potential to operate well within our organizations. However, we make decisions to fire people on the basis of their team-working skills, their collaborative styles and their approach to work. These are the skills that matter most to us as employers, and it is in these areas that employers have for many years looked to occupational psychologists for support. There are a large number of psychological profiling measures, most of which seek to provide a prose summary of the inter-personal styles of working likely to be adopted by an individual. These profile measures attempt to score, for example, the extent to which an individual might seek help, might use discussion and dialogue to move matters forward, or might be an effective solver of open-ended and ill-defined problems. SHL provide assessments such as OPQ and 16PF, which are conducted on-line and are widely used by employers. (See <http://www.shl.com/WhatWeDo/SHLReports/default.aspx> for the range of measures and reports that SHL provides.) The OPQ assessments seek to measure likely behaviors in three areas: Relationships with People; Thinking Style; and, Feeling and Emotions. For example, in measuring Feeling and Emotions, OPQ gauges the extent to which an individual is *relaxed*, *worrying*, *tough minded*, *optimistic*, *trusting*, and *emotionally controlled*. Similarly, OPQ measures a dimension called Influence and gauges the extent to which an individual is *persuasive*, *controlling*, *outspoken*, and *independent minded*. These – and other measures, such as Belbin's team styles – provide considerable overlap with the skills domain that interests 21st century educators and could well provide useful examples of the ways in which it is possible to assess students' ways of working.

## Tools for working

The newest set of skills is combined in this grouping of tools for working. These skills – information literacy and ICT literacy are the future and mark a major shift that is likely to be as important as the invention of the printing press. Friedman (2007), in fact, describes four stages in the growing importance of ICT. He identifies four "flatners" that are making it possible for individuals to compete, connect and collaborate in world markets.

- The introduction of personal computers that allowed anyone to author his/her own content in digital form that could then be manipulated and dispatched.
- The juxtaposition of the invention of the browser by Netscape that brought the internet to life resulting in the proliferation of websites and the overinvestment into fiber optic cable that has wired the world. In fact, NTT Japan has successfully tested a fiber optic cable that pushes 14 trillion bits per second that roughly equals 2,660 CD's or 20 million phone calls every second.
- The development of transmission protocols that made it possible for everyone's computer and software to be inter-operable. Consequently everyone could become a collaborator.

**Table 7: Tools for working - information literacy**

Knowledge	Skills	Attitudes/Values/Ethics
Access and evaluate information <ul style="list-style-type: none"> <li>• Access information efficiently (time) and effectively (sources)</li> <li>• Evaluate information critically and competently</li> </ul> Use and manage information <ul style="list-style-type: none"> <li>• Use information accurately and creatively for the issue or problem at hand</li> <li>• Manage the flow of information from a wide variety of sources</li> <li>• Apply a fundamental understanding of the ethical/legal issues surrounding the access and use of information</li> <li>• Basic understanding of the reliability and validity of the information available (accessibility/acceptability) and awareness of the need to respect ethical principles in the interactive use of IST.</li> </ul> Apply technology effectively <ul style="list-style-type: none"> <li>• Use technology as a tool to research, organize, evaluate and communicate information</li> <li>• Use digital technologies (computers, PDAs, media players, GPS, etc.), communication/networking tools and social networks appropriately to access, manage, integrate, evaluate and create information to successfully function in a knowledge economy.</li> </ul>	Access and evaluate information <ul style="list-style-type: none"> <li>• Ability to search, collect and process (create, organize, distinguish relevant from irrelevant, subjective from objective, real from virtual) electronic information, data and concepts and to use them in a systematic way;</li> </ul> Use and manage information <ul style="list-style-type: none"> <li>• Ability to use appropriate aids (presentations, graphs, charts, maps) to produce, present or understand complex information;</li> <li>• Ability to access and search a range of information media including the printed word, video and websites and to use internet-based services such as discussion fora and e-mail;</li> <li>• Ability to use information to support critical thinking, creativity and innovation in different contexts at home, leisure and work.</li> <li>• Ability to search, collect and process written information, data and concepts in order to use them in study and to organize knowledge in a systematic way. Ability to distinguish, in listening, speaking, reading and writing, relevant from irrelevant information.</li> </ul>	Access and evaluate information <ul style="list-style-type: none"> <li>• Propensity to use information to work autonomously and in teams; critical and reflective attitude in the assessment of available information.</li> </ul> Use and manage information <ul style="list-style-type: none"> <li>• Positive attitude and sensitivity to safe and responsible use of the Internet, including privacy issues and cultural differences.</li> <li>• Interest in using information to broaden horizons by taking part in communities and networks for cultural, social and professional purposes.</li> </ul>

- The expansion of the transmission protocols so that individuals could easily upload as well as download. For example, when the world was round, individuals could download vast amounts of information in digital formats that they could easily access and manipulate. But, in the flat world the key is the individual's ability to upload. This has given rise to open source courseware, blogs, and Wikipedia, to name only a few.

Just to paint a picture of how important it is to be truly literate in the use of these tools, consider that it is estimated that a week's worth of the New York Times contains more information than a person was likely to come across in a lifetime in the 18th century. And, it is estimated that four exabytes ( $4.0 \times 10^{19}$ ) of unique information will be generated this year – more than that the previous 5,000 years. In light of this information explosion the coming generations must have the skills to access and evaluate new information efficiently so that they can effectively utilize all that is available and relevant to their tasks at hand. One of the ways that they will manage this information explosion is through skilled use of ICT. Even now the use of ICT is growing. It has been reported that there are 31 billion searches on Google every month, up from 2.7 billion in 2006. To use Google one must effectively use the Internet. To accommodate the use of the Internet we have simultaneously seen an explosion in the number of Internet devices. In 1984, the number was 1,000, by 1992 it was 1,000,000, and in 2008 it had reached 1,000,000,000.





**Table 8: Tools for working - ICT literacy**

Knowledge	Skills	Attitudes/Values/Ethics
<p>Access and evaluate information &amp; communication technology</p> <ul style="list-style-type: none"> <li>• <i>Understanding of the main computer applications, including word processing, spreadsheets, databases, information storage and management;</i></li> <li>• <i>Awareness of the opportunities given by the use of Internet and communication via electronic media (e-mail, videoconferencing, other network tools); and the differences between the real and virtual world</i></li> </ul> <p>Analyze media</p> <ul style="list-style-type: none"> <li>• <i>Understand both how and why media messages are constructed, and for what purposes</i></li> <li>• <i>Examine how individuals interpret messages differently, how values and points of view are included or excluded, and how media can influence beliefs and behaviors</i></li> <li>• <i>Understand the ethical/legal issues surrounding the access and use of media</i></li> </ul> <p>Create media products</p> <ul style="list-style-type: none"> <li>• <i>Understand and know how to utilize the most appropriate media creation tools, characteristics and conventions</i></li> <li>• <i>Understand and know how to effectively utilize the most appropriate expressions and interpretations in diverse, multi-cultural environments</i></li> </ul>	<p>Access and evaluate information &amp; communication technology</p> <ul style="list-style-type: none"> <li>• <i>Access ICT efficiently (time) and effectively (sources)</i></li> <li>• <i>Evaluate information and ICT tools critically and competently</i></li> </ul> <p>Use and manage information</p> <ul style="list-style-type: none"> <li>• <i>Use ICT accurately and creatively for the issue or problem at hand</i></li> <li>• <i>Manage the flow of information from a wide variety of sources</i></li> <li>• <i>Apply a fundamental understanding of the ethical/legal issues surrounding the access and use of ICT and media</i></li> <li>• <i>Employ knowledge and skills in the application of ICT and media to communicate, interrogate, present and model</i></li> </ul> <p>Create media products</p> <ul style="list-style-type: none"> <li>• <i>Utilize the most appropriate media creation tools, characteristics and conventions, expressions and interpretations in diverse, multi-cultural environments</i></li> </ul> <p>Apply technology effectively</p> <ul style="list-style-type: none"> <li>• <i>Use technology as a tool to research, organize, evaluate and communicate information</i></li> <li>• <i>Use digital technologies (computers, PDAs, media players, GPS, etc.), communication/networking tools and social networks appropriately to access, manage, integrate, evaluate and create information to successfully function in a knowledge economy</i></li> <li>• <i>Apply a fundamental understanding of the ethical/legal issues surrounding the access and use of information technologies.</i></li> </ul>	<p>Access and evaluate information &amp; communication technology</p> <ul style="list-style-type: none"> <li>• <i>Be open to new ideas, information, tools and ways of working but evaluate information critically and competently</i></li> </ul> <p>Use and manage information</p> <ul style="list-style-type: none"> <li>• <i>Use information accurately and creatively for the issue or problem at hand respecting confidentiality, privacy and intellectual rights</i></li> <li>• <i>Manage the flow of information from a wide variety of sources with sensitivity and openness to cultural and social differences</i> <ul style="list-style-type: none"> <li>• <i>Examine how individuals interpret messages differently, how values and points of view are included or excluded, and how media can influence beliefs and behaviors</i></li> </ul> </li> </ul> <p>Apply and employ technology with honesty &amp; integrity</p> <ul style="list-style-type: none"> <li>• <i>Use technology as a tool to research, organize, evaluate and communicate information accurately and honestly with respect for sources and audience</i></li> <li>• <i>Apply a fundamental understanding of the ethical/legal issues surrounding the access and use of information technologies</i></li> </ul>

### *Information literacy*

Information literacy includes research on sources, evidence, biases, etc. Operational definitions are provided in. Operational definitions of information literacy are provided in Table 7. These are clearly increasingly important skills.

Category	Skills
Basic	Be able to open software, sort out and save information on the computer, and other simple skills in using the computer and software
Download	Be able to download different information types from the Internet
Search	Know about and how to get access to information
Navigate	Be able to orient oneself in digital networks, learning strategies in using the Internet
Classify	Be able to organize information according to a certain classification scheme or genre.
Integrate	Be able to compare and put together different types of information related to multimodal texts
Evaluate	Be able to check and evaluate if one has got the information one seeks to get from searching the Internet. Be able to judge the quality, relevance, objectivity and usefulness of the information one has found. Critical evaluation of sources.
Communicate	Be able to communicate information and express oneself through different mediational means.
Cooperate	Be able to take part in net-based interactions of learning, and take advantage of digital technology to cooperate and take part in networks
Create	Be able to produce and create different forms of information as multimodal texts, make web pages, and so forth. Be able to develop something new by using specific tools and software. Remixing different existing texts into something new.

### Box 2: Educational Testing Service framework for ICT literacy

The future consequences of recent developments in our societies due to globalization, networking (Castells, 1996) and the impact of ICT are spawning a set of new studies. See for example Hull and Schultz (2002) and Burbules and Silberman-Keller (2006) for examples on how such developments changes conceptions of formal and informal learning and what some term distributed or networked expertise (Hakkarainen, Palonen, Paavola & Lehtinen, 2004). Measurement procedures or indicators are still not clear with regard to these more future oriented skills. For example the *ImpaCT2* concept mapping data from the UK strongly suggests that there is a mismatch between conventional national tests, which focus on pre-specified knowledge and concepts, and the wider range of knowledge that students are acquiring by carrying out new kinds of activities with ICT at home (Somekh & Mavers, 2003). By using concept maps and children's drawings of computers in their everyday environments, the research generates strong indication of children's rich conceptualization of technology and its role in their world, for purposes of communication, entertainment, or accessing information. It shows that most children acquire practical skills in using computers that are not part of the assessment processes that they meet in schools. Some research has shown that students who are active computer users consistently under-perform on paper-based tests (Russell & Haney, 2000).

#### *ICT literacy*

EU countries, both on a regional and national level, and other countries around the world, are in the process of developing a framework and indicators to better grasp the impact of technology in education and what we should be looking for in assessing students' learning using ICT. For the EU see <http://europa.eu/rapid/pressReleasesAction.do?reference=IP/09/1244> and for Norway, Erstad (2006). For Australia, see Ainley et al. (2006). According to the Summit on Twenty-first Century Literacy in Berlin in 2002 (Clift, 2002), new approaches stress the abilities to use information and knowledge that extend beyond the traditional base of reading, writing, and math, which has been termed *digital literacy* or *ICT literacy*. Operational definitions of information literacy are provided in Table 8.

In 2001, the Educational Testing Service (ETS) in the U.S. assembled a panel for the purpose of developing a workable framework for ICT literacy. The outcome was the report *Digital transformation: A framework for ICT literacy* (International ICT Literacy Panel, 2002). Based on this framework, shown in Box 2, one can define ICT literacy as "the ability of individuals to use ICT appropriately to access, manage and evaluate information, develop new understandings, and

communicate with others in order to participate effectively in society” (Ainley et al., 2005). Different indicators of digital/ICT literacy can be (Erstad, in press).

In line with this perspective, some agencies have developed performance assessment tasks of “ICT Literacy,” indicating that ICT is changing our view on what is being assessed and how tasks are developed using different digital tools. One example is the tasks developed by the International Society for Technology in Education (ISTE) called *National Educational Technology Standards* (NETS, <http://cnets.iste.org/>), which are designed to assess how skilful students, teachers, and administrators are in using ICT.

In 2000 England’s Department for Education commissioned the development of an innovative test of 14 year-old students’ ICT skills. David Blunkett, at the time as Secretary of State for Education, described his vision for education and attainment in the 21st century. He spoke of raising expectations of student capabilities. He also announced the development of a new type of on-line test of ICT, which would assess the ICT skills students need in the 21st century. These assessments developed are outlined in Box 1 on p.11.

Development activity for the 14 year-olds test of ICT began in 2001. The original planned date for full roll-out and implementation was May 2009. In the event – and for a whole range of reasons – the original vision for the ICT tests was never realized. The test activities that were developed have been redesigned as stand-alone skills assessments that teachers in accredited schools can download and use informally to support their teacher assessment.

In Australia, a tool has been developed with a sample of students from grade 6 and grade 10 to validate and refine a progress map that identifies a progression of ICT literacy. The ICT literacy construct is described using three “strands”: working with information, creating and sharing information, and using ICT responsibly. Students carrying out authentic tasks in authentic contexts are seen as fundamental to the design of the Australian National ICT Literacy Assessment Instrument (Ainley et al., 2005). The instrument evaluates six key processes: accessing information (identifying information requirements and knowing how to find and retrieve information); managing information (organizing and storing information for retrieval and reuse); evaluating (reflecting on the processes used to design and construct ICT solutions and judgments regarding the integrity, relevance, and usefulness of information); developing new understandings (creating information and knowledge by synthesizing, adapting, applying, designing, inventing, or authoring); communicating (exchanging information by sharing knowledge and creating information products to suit the audience, the context, and the medium); and using ICT appropriately (critical, reflective and strategic ICT decisions, and considering social, legal, and ethical issues) (Ainley et al., 2005). Preliminary results of the use of the instrument show highly reliable estimates of ICT ability.

There are also cases where an ICT assessment framework is linked to specific frameworks for subject domains in schools. Reporting on the initial outline of a US project aiming at designing a Coordinated ICT Assessment Framework, Quellmalz and Kozma (2003) have developed a strategy to study ICT tools and skills as an integrated part of science and mathematics. The objective is to design innovative ICT performance assessments that could gather evidence of use of ICT strategies in science and mathematics.

## **Living in the world**

Borrowing the title of Bob Dylan’s song, to say that “the times they are a changin’” is a gross understatement when one considers how different living and working in the world will soon be. For example, the U.S. Department of Labor estimated that today’s learner will have ten to fourteen jobs by the age of 38. This reflects the rapidly growing job mobility with one in four workers having been with their current employer for less than a year, and one in two has been there less than five years. One might ask where these people are going as manufacturing and service industries move to places where there are abundant sources of cheap but sufficiently educated labor supplies. Essentially people must learn to live not only in their town or country but also in the world in its

**Table 9: Living in the world - citizenship, local and global**

Knowledge	Skills	Attitudes/Values/Ethics
<ul style="list-style-type: none"> <li>• <i>Knowledge of civil rights and the constitution of the home country, the scope of its government.</i></li> <li>• <i>Understand the roles and responsibilities of institutions relevant to the policy-making process at local, regional, national, and international level</i></li> <li>• <i>Knowledge of key figures in local and national governments; political parties and their policies.</i></li> <li>• <i>Understand concepts such as democracy, citizenship and the international declarations expressing them</i></li> <li>• <i>Knowledge of the main events, trends and agents of change in national, and world history</i></li> <li>• <i>Knowledge of the movements of peoples and cultures over time around the world</i></li> </ul>	<ul style="list-style-type: none"> <li>• <i>Participation in community /neighborhood activities as well as in decision-making at national and international levels; voting in elections.</i></li> <li>• <i>Ability to display solidarity by showing an interest in and helping to solve problems affecting the local or the wider community.</i></li> <li>• <i>Ability to interface effectively with institutions in the public domain.</i></li> <li>• <i>Ability to profit from the opportunities given by the home country and international programs.</i></li> </ul>	<ul style="list-style-type: none"> <li>• <i>Sense of belonging to one's locality, country, and (one's part of) the world.</i></li> <li>• <i>Willingness to participate in democratic decision-making at all levels.</i></li> <li>• <i>Disposition to volunteer and to participate in civic activities, support for social diversity and social cohesion.</i></li> <li>• <i>Readiness to respect the values and privacy of others with a propensity to react against anti-social behavior.</i></li> <li>• <i>Acceptance of the concept of human rights and equality; acceptance of equality between men and women.</i></li> <li>• <i>Appreciation and understanding of differences between value systems of different religious or ethnic groups.</i></li> <li>• <i>Critical reception of information from mass media.</i></li> </ul>

entirety. As more and more people individually move in the 21st century to compete, connect and collaborate, it is even more important that they understand all the aspects of citizenship. It is not enough to assume that what goes on in your own country is how it is or should be all over the globe. Hence, we have identified and group Citizenship, Life and Career, and Personal and Social Responsibility together as the 21st century skills.

### *Citizenship, global and local*

Citizenship as an educational objective is not new, and has been part of curricula, especially in social studies. A central focus has been on knowledge about democratic processes. Citizenship as a competence, however, has been growing in importance, and that implies certain challenges in measurement. In 2005, Operational definitions of citizenship are provided in Table 9.

Honey led a world-wide investigation into the use of 21st century assessments which investigated the existence and quality of assessments in key areas, including global awareness, concluding that "no measures currently exist that address students' understanding of global and international issues" (Ripley, 2007, p.5).

One example of a large-scale assessment of citizenship skills is the International Civic Education Study conducted by the International Association for the Evaluation of Educational Achievement (IEA). This research tested and surveyed nationally representative samples consisting of 90,000 14-year-old students in 28 countries, and 50,000 17 to 19-year-old students in 16 countries throughout 1999 and 2000.

The content domains covered in the instrument were identified through national case studies during 1996-1997 and included democracy, national identity, and social cohesion and diversity. The engagement of youth in civil society was also a focus. Torney-Purta, Lehmann, Oswald and Schulz (2001) reported the findings from these studies in the following terms.

- Students in most countries have an understanding of fundamental democratic values and institutions – but depth of understanding is a problem.
- Young people agree that good citizenship includes the obligation to vote.
- Students with the most civic knowledge are most likely to be open to participate in civic activities.
- Schools that model democratic practice are most effective in promoting civic knowledge and engagement.
- Aside from voting, students are skeptical about traditional forms of political engagement. But many are open to other types of involvement in civic life.
- Students are drawn to television as their source of news.
- Patterns of trust in government-related institutions vary widely among countries.
- Gender differences are minimal with regard to civic knowledge but substantial in some attitudes.
- Teachers recognize the importance of civic education in preparing young people for citizenship.

The main survey has been replicated as the International Civic and Citizenship Education Study in which data have been gathered in 2008 and 2009 and from which the international report is due for release in June 2010.

The developments of the Internet and Web 2.0 technologies have implications for the conception of citizenship as a competence. Jenkins (2006) says that these developments create a 'participatory culture'. This challenges, both locally and globally, the understanding of citizenship, empowerment and engagement as educational priorities. At the moment no measures exist which assess these skills in online environments, even though the research literature on 'young citizens online' has been growing in recent years (Loader, 2007).

One example of how these skills are made relevant in new ways is the Junior Summit online community. This consisted of 3,062 adolescents representing 139 countries. The online forum culminated in the election of 100 delegates. Results from one study indicate "the young online leaders do not adhere to adult leadership styles of contributing many ideas, sticking to task, and using powerful language. On the contrary, while the young people elected as delegates do contribute more, their linguistic style is likely to keep the goals and needs of the group as central, by referring to the group rather than to themselves and by synthesizing the posts of others rather than solely contributing their own ideas. Furthermore, both boy and girl leaders follow this pattern of interpersonal language use. These results reassure us that young people can be civically engaged and community minded, while indicating that these concepts themselves may change through contact with the next generation (Cassell, Huffaker, Ferriman & Tversky, 2006, p.436). In this sense it also relates to the German term 'Bildung' as an expression of how we use knowledge to act on our community and the world around us, that is what it means to be literate in a society, or what also might be described as cultural competence as part of broader personal and social responsibility.

### *Life and career*

The management of life and career is included among the skills needed for living in the world. There is long tradition of measurement of occupational preferences as one component for career guidance

**Table 10: Living in the world - life and career**

Knowledge	Skills	Attitudes/Values/Ethics
<p>Adapt to change</p> <ul style="list-style-type: none"> <li>• <i>Be aware that the 21st century is a period changing priorities in employment, opportunity and expectations</i></li> <li>• <i>Understand diverse views and beliefs particularly in multi-cultural environments</i></li> </ul> <p>Manage goals and time</p> <ul style="list-style-type: none"> <li>• <i>Understand models for long, medium and short term planning and balance tactical (short-term) and strategic (long-term) goals</i></li> </ul> <p>Be self-directed learners</p> <ul style="list-style-type: none"> <li>• <i>Identify and plan for personal and professional development over time and in response to change and opportunity</i></li> </ul> <p>Manage projects</p> <ul style="list-style-type: none"> <li>• <i>Set and meet goals, even in the face of obstacles and competing pressures</i></li> <li>• <i>Prioritize, plan and manage work to achieve the intended result.</i></li> </ul>	<p>Adapt to change</p> <ul style="list-style-type: none"> <li>• <i>Operate in varied roles, jobs responsibilities, schedules and contexts</i></li> </ul> <p>Be flexible</p> <ul style="list-style-type: none"> <li>• <i>Incorporate feedback effectively</i></li> <li>• <i>Negotiate and balance diverse views and beliefs to reach workable solutions</i></li> </ul> <p>Manage goals and time</p> <ul style="list-style-type: none"> <li>• <i>Set goals with tangible and intangible success criteria</i></li> <li>• <i>Balance tactical (short-term) and strategic (long-term) goals</i></li> <li>• <i>Utilize time and manage workload efficiently</i></li> </ul> <p>Work Independently</p> <ul style="list-style-type: none"> <li>• <i>Monitor, define, prioritize and complete tasks without direct oversight</i></li> </ul> <p>Interact effectively with others</p> <ul style="list-style-type: none"> <li>• <i>Know when it is appropriate to listen and when to speak</i></li> </ul> <p>Work effectively in diverse teams</p> <ul style="list-style-type: none"> <li>• <i>Leverage social and cultural differences to create new ideas and increase both innovation and quality of work</i></li> </ul> <p>Manage projects</p> <ul style="list-style-type: none"> <li>• <i>Set and meet goals, prioritize, plan and manage work to achieve the intended result even in the face of obstacles and competing pressures</i></li> </ul> <p>Guide and lead others</p> <ul style="list-style-type: none"> <li>• <i>Use interpersonal and problem solving skills to influence and guide others toward a goal</i></li> <li>• <i>Leverage strengths of others to accomplish a common goal</i></li> <li>• <i>Inspire others to reach their very best via example and selflessness</i></li> <li>• <i>Demonstrate integrity and ethical behavior in using influence and power</i></li> </ul>	<p>Adapt to change</p> <ul style="list-style-type: none"> <li>• <i>Be prepared to adapt to varied responsibilities, schedules and contexts, recognize and accept the strengths of others</i></li> <li>• <i>See opportunity ambiguity and changing priorities</i></li> </ul> <p>Be flexible</p> <ul style="list-style-type: none"> <li>• <i>Incorporate feedback and deal effectively with praise, setbacks and criticism</i></li> <li>• <i>Be willing to negotiate and balance diverse views to reach workable solutions</i></li> </ul> <p>Manage goals and time</p> <ul style="list-style-type: none"> <li>• <i>Accept uncertainty and responsibility and self manage</i></li> </ul> <p>Be self-directed learners</p> <ul style="list-style-type: none"> <li>• <i>Go beyond basic mastery to expand one's own learning</i></li> <li>• <i>Demonstrate initiative to advance to a professional level</i></li> <li>• <i>Demonstrate commitment to learning as a lifelong process</i></li> <li>• <i>Reflect critically on past experiences for progress</i></li> </ul> <p>Work effectively in diverse teams</p> <ul style="list-style-type: none"> <li>• <i>Conduct self in a respectable, professional manner</i></li> <li>• <i>Respect cultural differences, work effectively with people from varied backgrounds</i></li> <li>• <i>Respond open-mindedly to different ideas and values</i></li> </ul> <p>Produce results</p> <ul style="list-style-type: none"> <li>• <i>Demonstrate ability to:</i> <ul style="list-style-type: none"> <li>○ <i>Work positively and ethically</i></li> <li>○ <i>Manage time and projects effectively</i></li> <li>○ <i>Multi-task</i></li> <li>○ <i>Be reliable and punctual</i></li> <li>○ <i>Present oneself professionally and with proper etiquette</i></li> <li>○ <i>Collaborate and cooperate effectively with teams</i></li> <li>○ <i>Be accountable for results</i></li> </ul> </li> </ul> <p>Be responsible to others</p> <ul style="list-style-type: none"> <li>• <i>Act responsibly with the interests of the larger community in mind</i></li> </ul>

but no strong basis for building measures of skill in managing life and career. Suggestions for building operational definitions of this skill are provided in Table 10.

**Table 11: Living in the world - personal and social responsibility**

Knowledge	Skills	Attitudes/Values/Ethics
<ul style="list-style-type: none"> <li>• Know the codes of conduct and manners generally accepted or promoted in different societies.</li> <li>• Awareness of concepts of individual, group, society and culture and the historical evolution of these concepts.</li> <li>• Knowledge of how to maintain good health, hygiene and nutrition for oneself and one's family.</li> <li>• Knowledge of the intercultural dimension in their own and other societies</li> </ul>	<ul style="list-style-type: none"> <li>• Ability to communicate constructively in different social situations (tolerating the views and behavior of others; awareness of individual and collective responsibility).</li> <li>• Ability to create confidence and empathy in other individuals.</li> <li>• Ability to express one's frustration in a constructive way (control of aggression and violence or self-destructive patterns of behavior).</li> <li>• Ability to maintain a degree of separation between the professional and personal spheres of life, and to resist the transfer of professional conflict into personal domains.</li> <li>• Awareness and understanding of national cultural identity in interaction with the cultural identity of Europe and the rest of the world; ability to see and understand the different viewpoints caused by diversity and contribute one's own views constructively.</li> <li>• Ability to negotiate.</li> </ul>	<ul style="list-style-type: none"> <li>• Showing interest in and respect for others.</li> <li>• Willingness to overcome stereotypes and prejudices.</li> <li>• Disposition to compromise.</li> <li>• Integrity.</li> <li>• Assertiveness.</li> </ul>

### *Personal and social responsibility*

The exercise of personal and social responsibility is also included among the skills needed for living in the world. There are aspects of this skill in collaboration and teamwork, which is among the skills included among ways of working. Personal and social responsibility is taken to include cultural awareness and cultural competence. There is not a body of measurement literature on which to draw but the scope intended is set out in the operational definitions offered in Table 11.

## **Challenges**

The foregoing discussions have laid out principles for the assessment of 21st century skills, proposed 10 skills and given a sense of what they are and what measurements related to them might be built upon. That being said, there is still a very long row to hoe as it is not enough to keep perpetuating static tasks within the assessments. Rather, to reflect the need for imagination to compete, connect, and collaborate, it is essential that transformative assessments be created. This cannot begin to happen without addressing some very critical challenges.

This section summarizes key challenges to assessing 21st century skills in ways that truly probe the skills of students and provide actionable data to improve education and assessments.

### *Using models of skill development based on cognitive research*

The knowledge about acquisition of 21st century skills and their development is very limited, and the developers of assessments do not yet know how to create practical assessments using even this partial knowledge effectively (Bennet & Gitomer, 2009).

### *Transforming psychometrics to deal with new kinds of assessments*

Psychometric advances are needed to deal with a dynamic context and differentiated tasks, such as tasks embedded in simulations and using visualization that may yield a number of acceptable (and unanticipated) responses. While traditional assessments are designed to yield one right or best response, transformative assessments should be able to account for divergent responses, while measuring student performance in such a way that reliability of measures is ensured.

### *Making students' thinking visible*

Assessments should reveal the kinds of conceptual strategies a student uses to solve a problem. This involves not only considering students' responses, but also interpreting their behaviors that lead to these responses. Computers can log every keystroke made by a student and thus amass a huge amount of behavioral data. The challenge is to interpret the meaning of these data and link patterns of behavior to the quality of response. These associations could then illuminate students' thinking as they respond to various tasks.

That computers can score student responses to items effectively and efficiently is becoming a reality. This is certainly true of selected-response questions where there is a single right answer. It is also quite easy to apply partial credit models to selected-response items that have been designed to match theories of learning where not quite fully correct answers serve as the distracters. Constructed responses pose challenges for automated scoring.

The OECD's Program for the International Assessment of Adult Competencies (PIAAC) provides a good example of movement forward in machine scoring of short constructed responses. Some of the assessment tasks in PIAAC were drawn from the International Adult Literacy Survey (IALS) and the Adult Literacy and Life Skills (ALL) Survey where all answers were short constructed responses that needed to be coded by humans. By altering the response mode into either drop and drag or highlighting, the test developers converted the items into machine scoreable items. In these examples, however, all the information necessary to answer these types of questions resides totally in the test stimuli. Although the respondent might have to connect information across parts of the test stimuli, creation of knowledge not already provided is not required.

Machine scoring of extended constructed responses is in its infancy. Models do exist in single languages and are based on the recognition of semantic networks within responses. In experimental situations, these machine-scoring models are not only as reliable as human scorers but often achieve higher levels of consistency than can be achieved across human raters. Work has begun in earnest to expand these models to cross languages and may be available for international assessments in the foreseeable future.

### *Interpreting assisted performance*

New scoring rules are needed to take into account prompting or scaffolding that may be necessary for some students. Ensuring accessibility for as many students as possible and customization of items for special needs students within the design of the assessment is critical.



### *Assessing 21st century skills in traditional subjects*

Where the aims and goals of 21st century learning are described in countries' frameworks, they are generally specified as being taught through, within and across the subjects. However, computers can facilitate the creation of micro-worlds for students to explore in order to discover hidden rules or relationships. Tools such as computer-based simulations can in this way give a more nuanced understanding of what students know and can do than traditional testing methods. New approaches stress the abilities to use information and knowledge that extend beyond the traditional base of reading, writing, and mathematics. However, research shows that students still tuned into the old test situation with correct answers rather than explanations and reasoning skills can have problems in adjusting their strategies and skills. Without highly valued assessments of 21st century aims or goals requiring their teaching, it is difficult to see when or how education systems will change significantly for the majority of learners.

### *Accounting for new modes of communication*

To date newer modes of communication have rarely been represented in large-scale assessments. There is a mismatch between the skills young people gain in their everyday cultures outside of schools and the instruction and assessment they meet in schools. Different skills such as creativity, problem solving and critical thinking might be expressed in different ways using different modes and modalities, which ICT provides. In light of the developments described in the paper, it is essential that the radical changes in communication, including visual ways of communicating and social networking, be represented in some of the tasks of 21st century large-scale assessments. The speed with which new technologies develop suggest that it might be better to assess whether students are capable of rapidly mastering a new tool or medium than whether they can use current technologies.

### *Including collaboration and teamwork*

Traditional assessments are focused on measuring individual performance. Consequently, when faced with a collaborative task the most important question is how to assign credit to each member of the group, as well as how to account for differences across groups that may bias a given student's performance. This issue arises whether students are asked to work in pre-assigned complementary roles or whether they are also being assessed on their skills in inventing ways to collaborate in an undefined situation. Questions on assigning individual performance as well as group ratings become even more salient for international assessments where cultural boundaries are crossed.

### *Including local and global citizenship*

The assessment of citizenship, empowerment and engagement, both locally and globally, is underdeveloped. At this time, no measures exist that assess these skills in online environments, even though the research literature on 'young citizens online' has been growing in recent years. For international assessments, cultural differences and sensitivities will add to the challenge of developing tasks valid across countries. Having students solve problems from multiple perspectives is one way to address the challenge of cultural differences.

### *Ensuring validity and accessibility*

It is important to ensure validity of standards on which assessments are based; accessibility with respect to skills demands, content pre-requisites and familiarity with media or technology; and an appropriate balance of content and intellectual demands of tasks.

These important attributes of any assessments will prove particularly challenging for the transformative assessments envisaged in this paper. Careful development and piloting of innovative

tasks will be required, including scoring systems that ensure comparability of complex tasks. Fluidity studies with technology are important in devising tasks for which experience with technology does not predict performance. Also, complex tasks typically demand access to intellectual resources (e.g., a search engine); therefore, this needs to be factored into designing complex assessment tasks as envisaged for transformative assessments.

*Considering cost and feasibility*

Cost and feasibility are factors operating for any assessment but will be greatly exacerbated for the innovative and transformative assessments that are to address the kinds of 21st century skills discussed in this paper. For sophisticated online assessments, ensuring that schools have both the technical infrastructure needed and the controls for integrity of data collection is mandatory. These latter matters are considered in White Paper 3.

## White Paper 1 Annex: Expanded descriptions of assessment tasks

### UK ICT Test for 14-year-old students

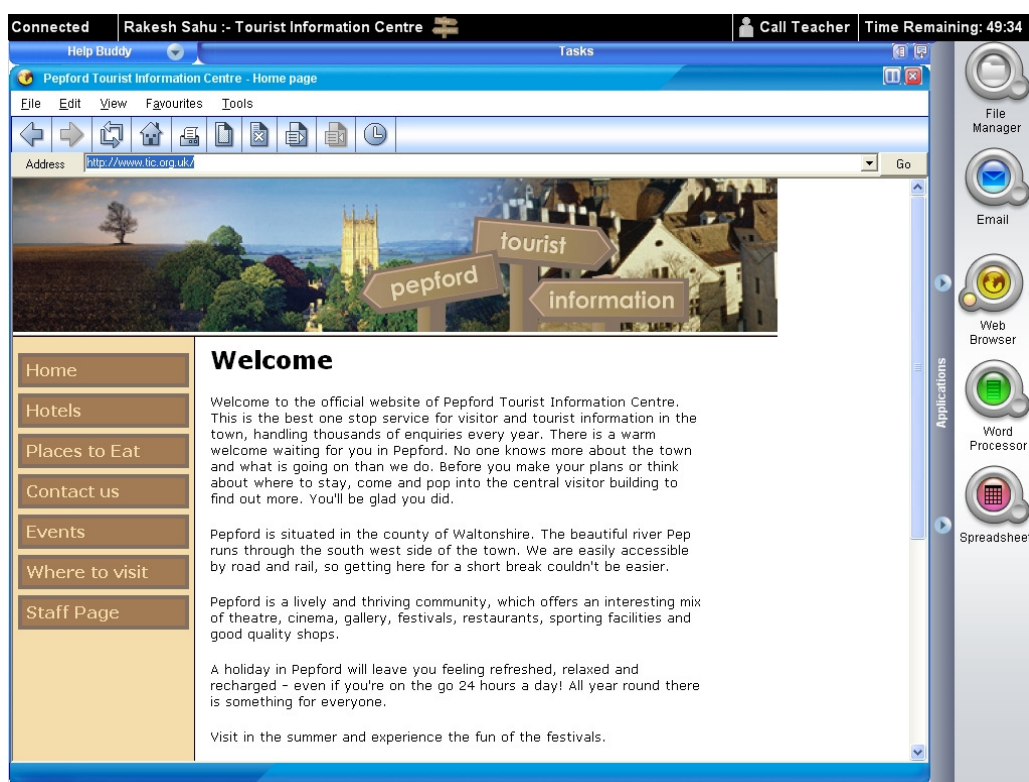
Publisher:	Qualifications and Curriculum Authority (QCA) and the UK Education Department, currently the Department for Children, Schools and Families (DCSF).
Contact:	QCA and/or Martin Ripley martin.ripley1@btinternet.com
<b>Brief details</b>	
<p>In 2000 England's Department for Education commissioned the development of an innovative test of 14 year-old students' ICT skills. David Blunkett, at the time as Secretary of State for Education, described his vision for education and attainment in the 21<sup>st</sup> century. He spoke of raising expectations of student capabilities. He also announced the development of a new type of on-line test of ICT, which would assess the ICT skills students need in the 21<sup>st</sup> century. Development activity for the 14 year-olds test of ICT began in 2001. The original planned date for full roll-out and implementation was May 2009. In the event - and for a whole range of reasons - the original vision for the ICT tests was never realised. What is described in this paper is based on the early vision for the ICT tests, which guided development activity until 2004.</p> <p>The ICT tests are no longer being developed in the UK. The test activities that were developed have been redesigned as stand-alone, closed item skills assessments, which teachers in accredited schools can download and use informally to support their teacher assessment.</p>	
<b>The innovation</b>	
<p>The ICT test development project was designed to create an on screen assessment which would change the nature of testing. It was designed to:</p> <ul style="list-style-type: none"> <li>• Create a simulated, virtual-world environment within which students would complete tests</li> <li>• Create a desktop environment with software and tools for students</li> <li>• Create new ways of scoring student performances, based on evidence of the ICT processes they use to solve problems (rather than scoring the student's ICT product)</li> <li>• Create new ways of enabling access to tests for all students</li> <li>• Be accompanied by the training and support infrastructure needed by teachers</li> <li>• Stimulate the development of hardware environments within schools capable of sustaining this form of testing</li> </ul>	
<b>Description</b>	
<p>The ICT tests were based on the assessment of four ICT capabilities:</p> <ol style="list-style-type: none"> <li>1. Finding things out. For example, <i>how to obtain information well matched to purpose by selecting appropriate sources; or, questioning the plausibility and value of information found.</i></li> <li>2. Developing ideas and making things happen. For example, <i>use ICT to measure, record, respond to and control events.</i></li> <li>3. Exchanging and sharing information. For example, <i>use ICT to share and exchange information, such as web publishing and video conferencing.</i></li> <li>4. Reviewing, modifying and evaluating work as it progresses. For example, <i>reflect critically on their own and others' use of ICT.</i></li> </ol>	

The tests were designed to take place in a “walled garden”. Students taking the tests logged into simulated ICT environment. This environment was designed to mimic real-world ICT and internet environments, and to reflect human and business uses of ICT. The environment had two components:

### A Virtual World

This Virtual World centred around the creation of a simulated world - called Pepford. Pepford consisted of a wide range of web-sites, reflecting the range of resources and information (and mis-information!) that might be found in reality. For example, these covered bus and train companies, leisure centres, schools, weather stations and so on. Some aspects of these websites could be configured to adapt dynamically. Some of these adaptations were designed to add authenticity - such as updating dates, colours and cosmetic features. Other aspects of the sites could be configured to adapt according to the performance of the student.

One example screen shot is set out below. This shows the welcome screen that greeted students as they first logged into the test environment.

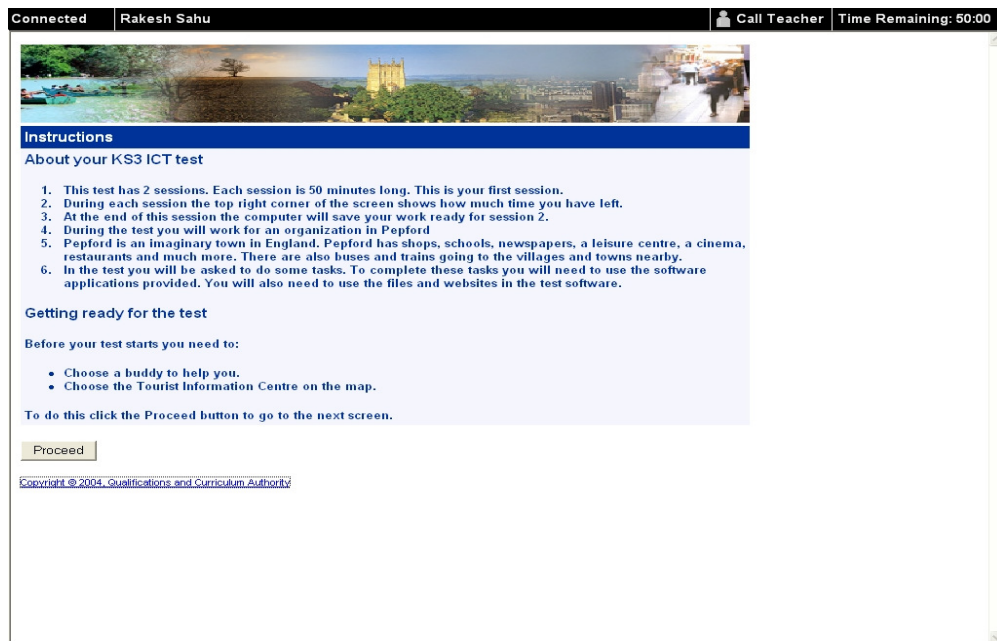


### A Virtual Desktop Environment

A virtual desktop environment was also created for the ICT tests. The purpose of this virtual environment was to provide a common set of applications to all students (rather than students taking the tests on their own schools' desktop environments, which would vary significantly). The functionality of the applications was researched to ensure that it would be similar to, and therefore reasonably intuitive, the desktops commonly in use in schools. Initially, applications were developed covering: email, file management, web browser, word processing, spreadsheet, database and control.

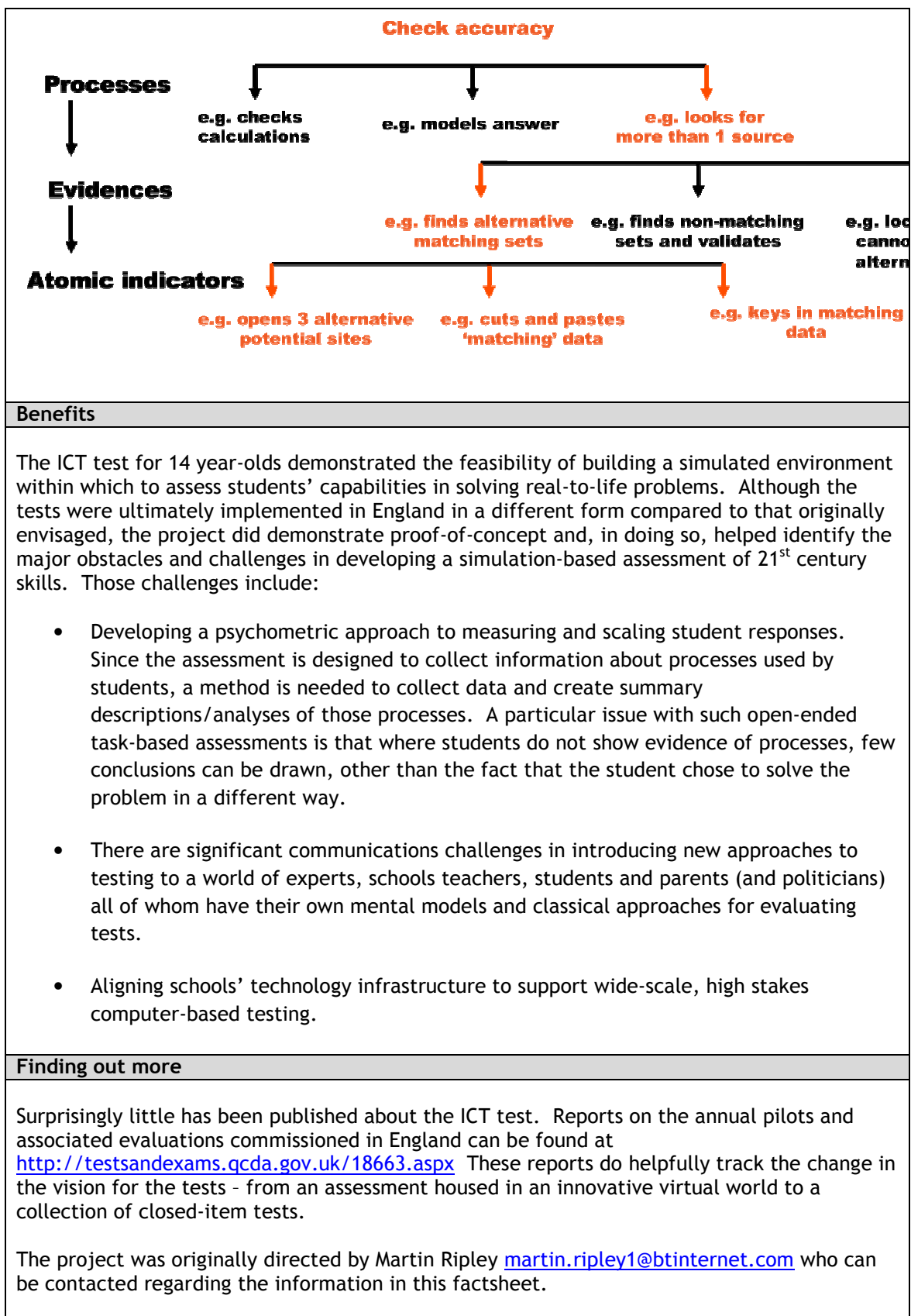
### Assessment Activities

Students complete two 50-minute test sessions. In these two sessions a student is assigned a task to complete. When a student successfully logs into a test they would be taken to the following landing page, which contains confirmatory instructions regarding the test session and the task.




A wide range of authentic tasks were designed for students to complete in the two test sessions. For example, as students log into the virtual world of Pepford, they might receive an email ostensibly from the editor of one of the local Pepford on-line news-sites, asking the student to research job vacancies in Pepford and to prepare a vacancies page for the web-site. To complete this task, students would need to research current vacancies in Pepford by running web-searches and by emailing virtual companies to request more information about vacancies. The extent and quality of information available would vary - some sites would contain comprehensive lists of vacancies, others would provide the email address of a person to contact. Some sites would contain current updates; others might appear older or even out of date. This variation and complexity was intended to reflect real-world ICT information environments. During the course of completing the task, a student would typically receive further requests from the editor, perhaps changing deadlines or adding requirements - just as happens in the real world.

A student's work is graded automatically. The assessment is based on the collection of evidence of the ICT processes that students complete. For example, the ICT capability of "check accuracy" could be evidenced by a student through a number of processes, including "looks for more than 1 source". In turn, this process could be evidenced in a number of ways, including "finds alternative matching sets" which students would achieve through a series of key-board or mouse actions (atomic indicators). This automated scoring of students' work was determined through the creation of a complex Rules Base which was created through empirical observation and expert review of student activity in completing the tasks. An example of the Rules Base, based on the above example of "check accuracy", is set out below.



**eSCAPE**

<b>Publisher:</b>	The Qualifications and Curriculum Authority (QCA) and Becta.
<b>Contact:</b>	Professor Richard Kimbell, TERU, Goldsmiths College, The University of London, UK
<b>Brief details</b>	
<p>For many years, England's school examinations for 16 year-old students have included an optional assessment in Design and Technology. Traditionally this examination includes a requirement for students to complete a design project of over 100 hours duration, and for the student to submit a written report on the project. The report is then graded.</p> <p>In 2003, the Qualifications and Curriculum Authority commissioned the Technology Education Research Unit (TERU) at Goldsmiths College in London to undertake to develop a technology-led replacement to this traditional paper-based assessment. The result is an assessment completed in 6 hours, in a design workshop, with students working in groups of 3 or 4. During the course of the 6 hours, students are given a number of staged assessment instructions and information via a personal, portable device. The handheld device also acts as the tool to capture assessment evidence - via video, camera, voice, sketch pad and keyboard. During the 6 hours, each student's design prototype develops, with the handheld device providing a record of progress, interactions and self-reflections.</p> <p>At the end of the assessment, the assessment evidence is collated into a short multi-media portfolio. This is viewed by human raters, who score each student's responses. In order to provide an holistic judgement on the students' work, eSCAPE directors turned to the work of Thurstone to develop a graded-pairs scoring engine. This engine supports human raters in making a number of paired judgements about students' work. The result is an assessment which exhibits rates of reliability equal to, or slightly in excess of, the levels of reliability achieved on multiple-choice tests.</p> <p>Project eSCAPE is ongoing.</p>	
<b>The innovation</b>	
 <p>Project eSCAPE was established to develop innovative solutions to three problems:</p>	

- Finding a technology solution, with accompanying software, enabling students to capture a dynamic record of their activity in responding to an assessment of their creativity and design capability
- Building a assessment model which would provide a more valid assessment of creativity and design than paper-based counterparts
- Finding or designing a measurement model which would preserve the integrity of the assessment whilst also delivering outcomes sufficiently reliable within a high stakes testing environment.

#### Description



#### *Building a prototype system*

The initial phase of Project eSCAPE concluded with an analysis of the **specification of a prototype system** that could theoretically facilitate the creation of real-time dynamic portfolios direct from the workshop/classroom.

Subsequently the project directors sought to construct a *working prototype* that could be used in a school pilot with approximately 20 schools. eSCAPE phase 2 involved building and testing a prototype system in which learners undertake a design coursework activity in the classroom (using hand-held digital tools) and their resulting portfolio emerges automatically (and dynamically as they work) in a website where it can be displayed for assessment purposes.

Design studios and workshops are typically not full of desk-top or lap-top computers - which are often down the corridor in a different room - often an IT suite. Since we were concerned to explore the **dynamic creation** of an e-portfolio (rather than a sanitised one created through 2<sup>nd</sup>



hand re-telling of the real story) we chose to use peripheral digital tools (eg digital pens, cameras, PDAs) that were capable of being used in the workshop/studio setting.

This phase of the project was focused on two things:

- technical system development to enable the dynamic links between learners (doing the activity) , teachers (running the activity) , web-servers (managing and storing the communications), and the portfolio repository (to display the resulting portfolios).
- developing appropriate assessment activities along with all the associated resources to make them work effectively in design studios and workshops throughout the country.

The project focused on running the pilot in 11 schools during June/July 2006, and accumulated 250 portfolios in the website. The assessment of these portfolios involved the development of a completely new measurement methodology that was needed because of the web-based nature of the work. The project utilised a 'differentiated pairs' methodology (Thurstone 1927), enabling a small team of judges to assess the students' work.

The outcomes of eSCAPE phase 2 were essentially of three kinds:

- i) concerning the classroom activity, it was established that eSCAPE activities of this kind could operate successfully in school workshops and studios. In every school the activity ran successfully and in the process it became evident that learners were very enthusiastic about operating in this way as part of their normal project-work.
- ii) concerning the technology, it was established that the technology underpinning the eSCAPE system was reliable and manageable. The system operated from hand-held devices in the classroom through secure web-servers and learners' portfolios successfully emerged in the web-space.
- iii) concerning the assessment process, it was established that the comparative pairs methodology operated very effectively. It proved to be very manageable at the judging end of the process, and it yielded astonishingly reliable results at the outcome.

During this phase of eSCAPE there was significant interest from the assessment and teaching community. Three videos were created by TeachersTV, and these can be viewed at the following links.

<http://www.teachers.tv/video/3306>

<http://www.teachers.tv/video/3307>

<http://www.teachers.tv/video/5431>

### *Scoring students' work*

Assessing web-based portfolios can be done using the same systems as are conventionally used for assessing paper-based work, by allocating scores for elements within a rubric. But having all the work in a website opens the possibility of using a quite different approach. Alastair Pollitt worked with the TERU team to create an approach of 'comparative pairs' judgements that was developed originally from the work of Thurstone (1927) in the 1920s. This involves a judge looking at two portfolios and deciding (only) which - on the basis of agreed criteria - is the better / stronger piece of work. Then looking at another pair, and another pair, and another pair. Many paired judgements are made, enabling each piece of work to be compared to many other pieces, and the process is undertaken by a team of judges so that each piece is seen by many different judges. The combined effect of all this is two-fold.

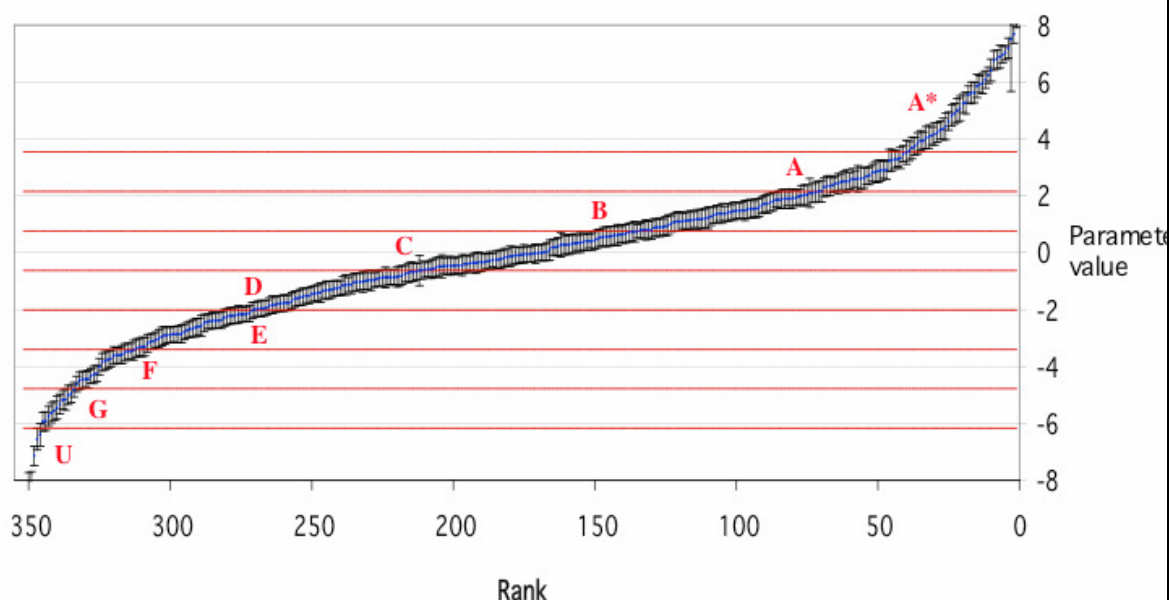
First, using Rasch analysis, the mass of wins and losses for individual portfolios can be transformed into a rank order of all of the portfolios. Those at the top are those that win every comparison and those at the bottom have lost every time. In the middle are those that win half the time and lose half the time.

Second, since in our case each portfolio was judged at least 17 times (sometimes significantly more) and by 7 judges, the judging process renders highly reliable results. The standard error attaching to the placement of individuals within the rank order is significantly lower than would be the case in conventional portfolio assessment.

The judging process (including training of judges and 3 rounds of judging) was undertaken in September and October 2006 and the resulting data used for analysis of learners' performance. Whilst the pilot was principally designed to test the system itself, it was necessary to test learners within the system and the resulting data has proved interesting concerning for example the relationships in performance between designing on paper and designing digitally; between

girls and boys performance; and in relation to different 'general ability' groups. Overall however, we were concerned to gauge teachers' and learners' reaction to the activity system in schools and the judges' reaction to the assessment system. A further pilot of the Thurstone approach was conducted in 2008. This again provided reliability coefficients of 0.94.

Parameters, Standard Errors and 'Grades'



### Benefits

Project eSCAPE has developed solution to four broad categories of research questions.

#### Technology

The eSCAPE system is driven by a remote server dynamically sending and receiving data to and from hand-held digital tools, putting the teacher in control of the sequences of the task and automatically building an evidence trail in the web portfolio. The mobile, digital technology has integrated well - almost invisibly - into the design workshop/classroom.

#### Pedagogy

Teachers have consistently rated the eSCAPE approach very positively, supporting the view that the assessment enhances classroom teaching and learning in the areas of design, creativity and collaboration. The approach has now also won the attention and support of geography and science assessment specialists.

#### Functionality

Assessment of student performances is notoriously difficult. It is difficult to manage the performance itself in ways that assure equity to all learners and it is difficult to ensure reliability in the assessment. eSCAPE systems work well in supporting performance, and the data shows that it produced highly reliable assessment statistics.

#### **Finding out more**

Professor Richard Kimbell is the Director of project eSCAPE and the author of numerous reports and presentations on eSCAPE. Much of the above text has been taken from his report. A great deal of information about the project is published on the TERU website at

<http://www.gold.ac.uk/teru/projectinfo/projecttitle,5882,en.php>

The latest 2008-09 project report will shortly be published also.

The following links are to video broadcasts produced by TeachersTV and which provided a detailed narrative of eSACPE.

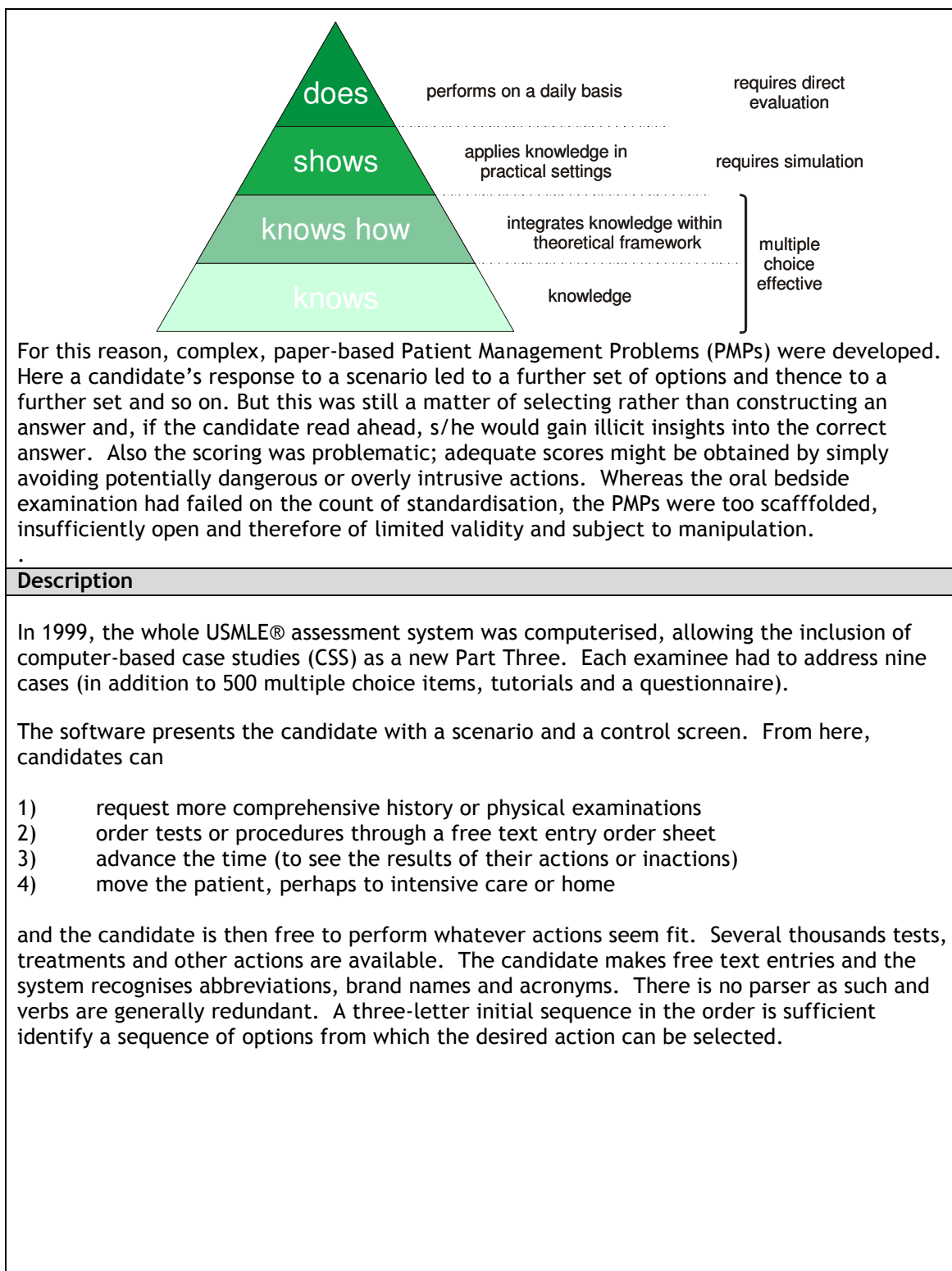
<http://www.teachers.tv/video/3306>

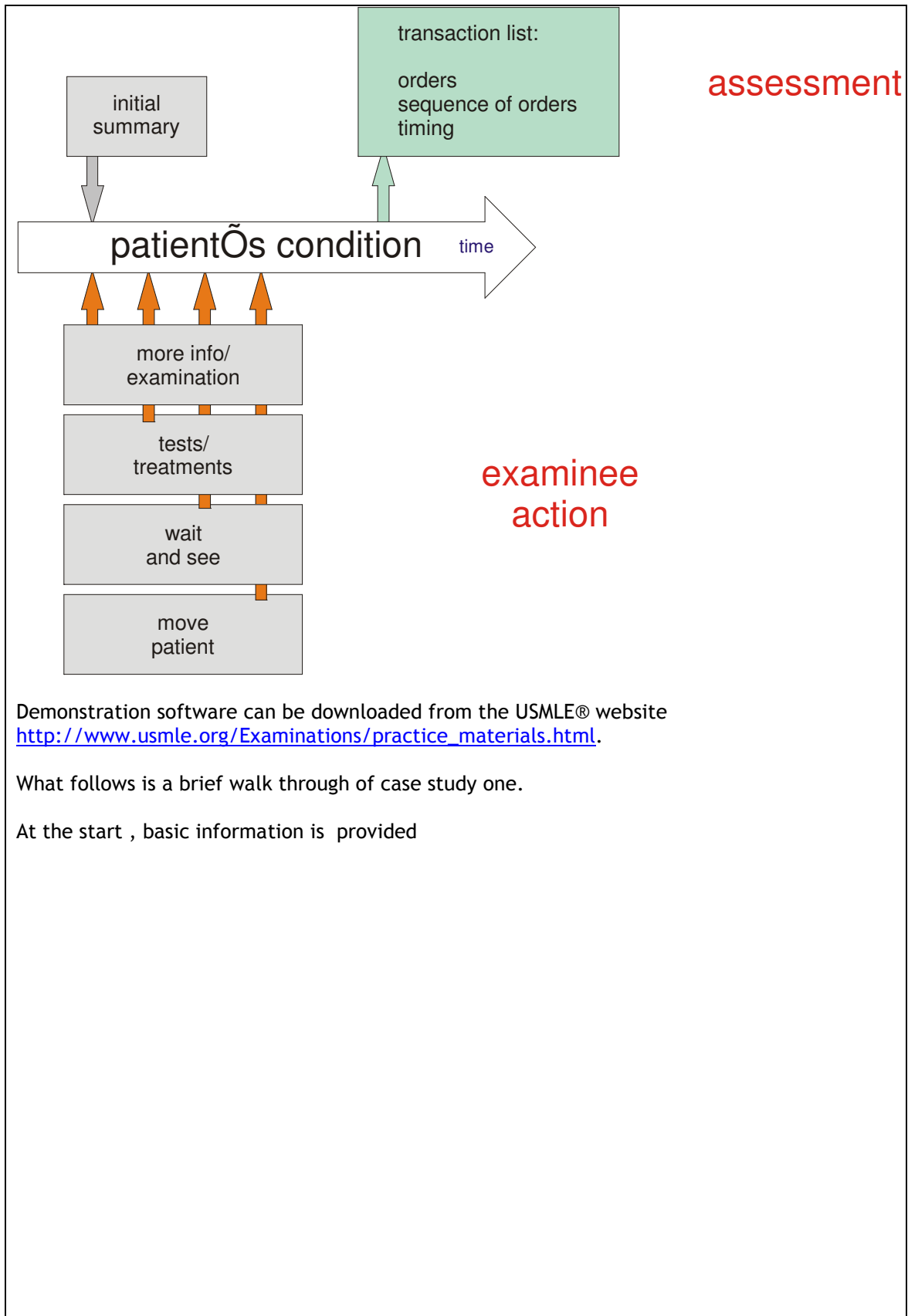
<http://www.teachers.tv/video/3307>

<http://www.teachers.tv/video/5431>

**Primum® Computer-based Case Simulations (CCS) for licensing doctors**

<b>Publisher:</b>	National Board of Medical Examiners (NBME)
<b>Contact:</b>	National Board of Medical Examiners (NBME) 3750 Market Street Philadelphia, PA 19104-3102
<b>Brief details</b>	
<p>Primum® is used as one of three parts of the United States Medical Licensing Examination (USMLE®) which is designed to test would-be doctors' ability to treat patients in a practical setting, skills that were previously examined at the bedside. It is a high-stakes, computer-based case study simulation where candidates are presented with authentic problems and are asked to treat a simulated patient on screen. Candidates, using free text entry, receive information, conduct examinations and order tests and treatments, to which the electronic patient will respond. A candidate's performance is assessed against model responses using a regression-based, automated scoring procedure. The technique is powerful and effective and would appear to be relevant to other subject areas/professional training courses where "doing skills" are important and systems can be modelled eg economics.</p>	
<b>The innovation</b>	
<p>The United States Medical Licensing Examination™ (USMLE™) is the present three-step examination for medical licensure in the United States and is sponsored by the Federation of State Medical Boards (FSMB) and the National Board of Medical Examiners® (NBME). They have been developed from the earlier NBME examinations. In these, until the early 1950s, essays and oral examination were predominant.</p> <p><u>1922- 1950s</u></p> <p>Part 1 3 day essay examination in the basic sciences at completion of 2nd year of medical school</p> <p>Part 2 2 day written examination in the clinical sciences at graduation from medical school</p> <p>Part 3 1 day practical oral examination on clinical and laboratory problems conducted at the bedside at end 1st year post graduate.</p> <p>In the early 1950s, a NBME and Education Testing Services (ETS) study explored the potential advantages of replacing essays with multiple choice tests and concluded that the latter offered greater reliability and validity, conforming more closely with instructor judgements. They were adopted for Parts One and Two. Part Three remained the same.</p> <p>In the 1960s, concerns about inter-rater reliability and resource implications in terms of both examiners and patients led to the dropping of the Part Three bedside oral component. But it was felt that multiple choice questions reliably addressed lower taxonomic levels but not higher ones and that it was unsatisfactory to rely solely on multiple choice items for the assessment of clinical skills.</p>	





The screenshot shows a medical simulation interface. At the top, there are four callout boxes with red text: "How are you?", "Order tests, procedures", "Advance time", and "Move patient eg to ICU, home". Below these is a toolbar with four icons: a stethoscope and pulse (Interval Hx or PE), a clipboard (Write Orders or Review Chart), a clock (Obtain Results or See Patient Later), and a red cross (Change Location). The toolbar also displays "Day 1 @ 16:00 (Wed)" and "Emergency department". The main window is titled "Case Introduction" and shows the text: "Day 1 @ 16:00", "Emergency department", and "A 65-year-old white man is brought to the emergency department because of sharp chest pain and respiratory distress. He is in acute distress, moaning, and holding his hands over the right side of his chest." A callout box with red text "initial scenario" points to the text. An "OK" button is at the bottom of the window. At the very bottom of the interface, a status bar reads: "lapsed SIMULATED Case Time = 0 Days 0 Hrs 0 Min | Elapsed REAL time < 1 minute".

You (the candidate) call up the initial vital signs and a slightly fuller history.

The screenshot shows a medical simulation interface with a window titled "Initial vital signs". The window displays the following information: "Day 1 @ 16:00", "Initial vital signs", "Temperature: 37.0 degrees C (98.6 degrees F)", "Pulse: 120 beats/min", "Weak", "Respiratory rate: 34 /minute", "Blood pressure, systolic: 100 mm Hg", "Blood pressure, diastolic: 60 mm Hg", "Height: 183 cm (72.0 in)", "Weight: 97.5 kg (215.0 lb)", and "Body mass index: 29.1 kg/m2". An "OK" button is at the bottom of the window.



**Initial history**

**Initial history**  
Reason(s) for Visit:  
Chest pain; respiratory distress

**History of Present Illness:**  
The patient, a 65-year-old accountant, is brought to the emergency department by ambulance from the trucking company where he works. About 10 minutes before the ambulance arrived, the patient developed excruciating, sharp pain in the right side of his chest and marked respiratory distress. He rates the pain as 8 on a 10-point scale. The pain increases with respiration. He is unable to answer questions. A coworker who accompanied the patient to the hospital says that this never happened before, but the patient has had emphysema and asthma for years. Oxygen was administered during transport.

All other history unobtainable.

OK

You (the candidate) call up a physical examination:

Help

Interval Hx or PE    Write Orders or Review Chart    Obtain Results or See Patient Later    Change Location

Day 1 @ 16:00 (Wed)    Emergency department

**History**

☐ Interval/follow up

**Physical Examination**

☐ Complete  
(Order VS from order sheet)

☐ General Appearance  
☐ Skin  
☐ Breasts  
☐ Lymph Nodes  
☐ HEENT/Neck  
☒ Chest/Lungs

☒ Heart/Cardiovascular  
☐ Abdomen  
☐ Genitalia  
☐ Rectal  
☐ Extremities/Spine  
☐ Neuro/Psych

Problem appears to be pulmonary – order chest/lungs/ cardiovascular examination

OK    Clear

Lapsed SIMULATED Case Time = 0 Days 0 Hrs 0 Min | Elapsed REAL Time = 2 minutes

... and check the results.

History and Physical

Day 1 @ 16:02

**Chest/Lung examination**  
Chest wall normal. Hyperresonance to percussion on right; no breath sounds. Breath sounds present on left.

**Cardiovascular examination**  
Tachycardia; heart sounds faint. No murmurs, rubs, gallops, or extra sounds. Bilateral central and peripheral pulses weak but equal. No jugular venous distention.

results of examination

OK

You decide to order cardiac monitoring, and pulse oximetry to assess oxygen saturation. As soon as the absent breath sounds are discovered, you order a needle thoracostomy followed by a chest tube insertion.

Help

Interval Hx or PE
Write Orders or Review Chart
Obtain Results or See Patient Later
Change Location
Emergency department

ORDER SHEET | Progress

Order
Monitor, cardiac
Pulse oximetry

Order Verification
nee
Needle aspirate, peritoneal abscess, without imaging guidance
Needle aspirate, peritoneal abscess, with imaging guidance
Needle aspirate, prostate
Needle aspirate, subdiaphragmatic abscess, without imaging guidance
Needle aspirate, subdiaphragmatic abscess, with imaging guidance
Needle aspirate, subhepatic abscess, without imaging guidance
Needle aspirate, subhepatic abscess, with imaging guidance
Needle aspirate, subphrenic abscess, without imaging guidance
Needle aspirate, subphrenic abscess, with imaging guidance
Needle biopsy, bone, imaging guided
Needle biopsy, bone marrow
Needle biopsy, kidney
Needle biopsy, liver
Needle biopsy, lung
Needle biopsy, pancreas
Needle biopsy, thyroid
Needle drainage, abscess, imaging guided
Needle drainage, abscess, nonimaging guided
Needle drainage, gallbladder, imaging guided
Needle precautions
Needle thoracentesis
Needle thoracostomy

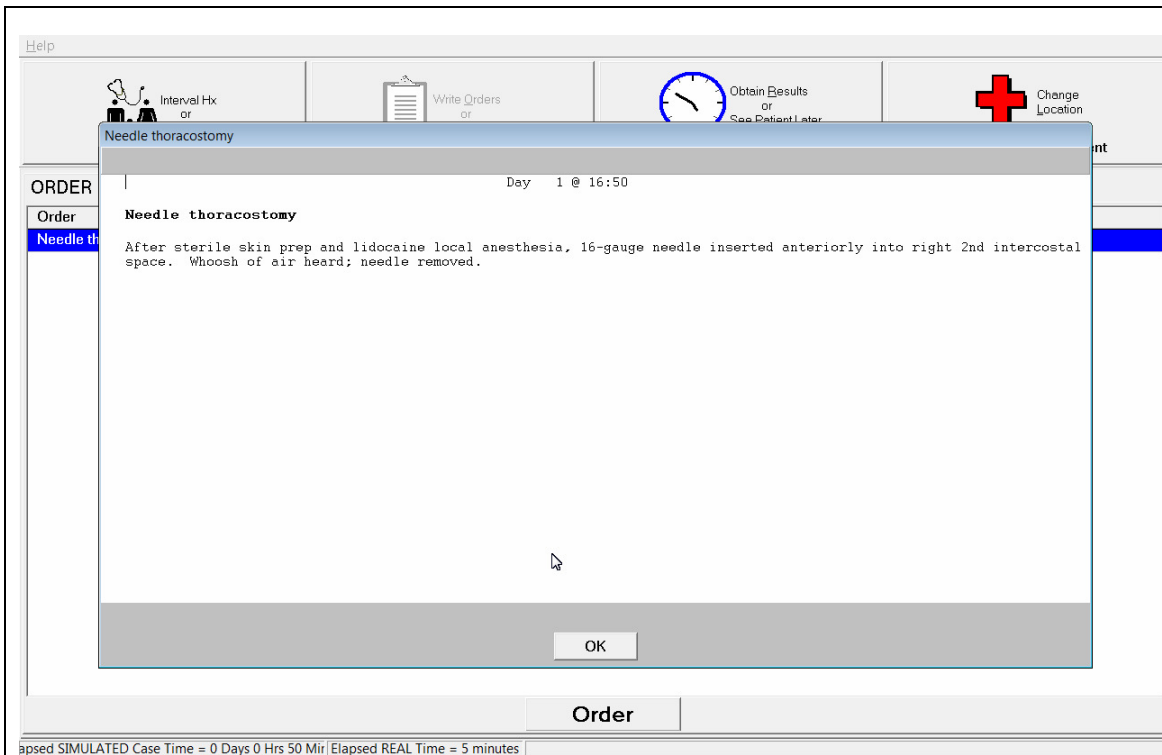
typing 'nee' brings up this screen. Select 'needle thoracostomy'

OK Cancel Broaden Search

Order

lapsed SIMULATED Case Time = 0 Days 0 Hrs 6 Min | Elapsed REAL Time = 7 minutes

A chest x-ray would need to be ordered to make sure the tube was inserted in the right place and the blood pressure and respiratory rate should be monitored until the patient's condition has stabilized. The effect of any actions you order will be revealed.



At the end of the case study you are asked for a diagnosis and thanked for looking after the patient (should your treatment be less than optimal, the cases end before the patient actually dies to avoid affecting the examinee in a way that could impact their performance on the next case). The system has stored a transaction list of all your actions, their sequence and timing. It is this transaction list which is scored. (In addition the system stores a complete record of keystrokes that can be used for research purposes.)

### designing a scenario

Primum® uses proprietary software developed by the NBME. The pathways associated with each patient scenario are developed separately and are complex. The number of possibilities can be controlled either by having the case end or by having the patient refuse certain treatments. So for example, the patient might refuse completely inappropriate (and essentially unanticipated) surgery. That requested surgery counts against the examinee in scoring, but doesn't change the patient's condition because it never happened.

Creating each scenario is time consuming and expensive. To be cost effective scenarios must be kept secure and reused.

### scoring and the linear regression model

Experts have to consider the effects of any of thousands of actions and determine whether they are beneficial, neutral or dangerous. Beneficial and dangerous actions need to be rated for degree - eg essential, important or desirable. The appropriateness of actions will also be dependent on timing and sequence. A numerical score is generated based on a linear regression model to produce a score which an expert would have produced. This is done as follows.

Using the software, experts explore the scenario and produce a model answer (including

actions and timings) and the associated mark. They then specify beneficial and dangerous actions and associate them with score bands.

These ratings are then tested by the experts, by independently marking sample transaction lists generated by examinees, and discussing their scores. Through an iterative process they achieve common understanding (if not consensus), their scores are averaged and the mean rating used as the dependent measure in a regression equation.

### Benefits

Authentic testing was becoming too expensive, too resource-hungry and did not deliver standardised results. The CCS model presents potential doctors with authentic problems where they have to manage the patient in a realistic way. The linear regression model allows judgments to be more consistent than the level achieved through the direct use of experts. The procedure has been developed over several years and demands continuing appreciable high levels of resourcing in terms of expert panel input. However, it shows considerable resource savings over direct bedside assessment as well as an appreciable increase in standardisation (reliability).

Unlike the PMPs, the computer retains a complete listing of what the candidate has done while managing the simulated case.

It is significant that time is a factor in assessing potential doctors. Where this is not the case, other approaches can be used eg rules based methods (see below).

### Finding out more

Download the *Primum® Computer-Based Case Simulations (CCS) - Tutorial and Practice Cases* at <http://www.usmle.org/Orientation/2009/menu.html>

More information is available at [www.usmle.org](http://www.usmle.org)  
[www.usmle.org/Examinations/step3/step3.html](http://www.usmle.org/Examinations/step3/step3.html)  
[www.usmle.org/Examinations/practice\\_materials.html](http://www.usmle.org/Examinations/practice_materials.html)

More details of the theory behind Primum® are in *Automated Scoring of complex tasks in Computer-Based Testing*, Williamson Mislevy and Bejar; Mahwah New Jersey 2006, chapter A regression-based procedure for automated scoring of a complex medical performance assessment, Melissa J. Margolis and Brian E. Clauser.

More details on rule-based methods are in *Williamson Mislevy and Bejar op cit*, chapter Henry Braun et al, Rule-based methods for automatic scoring: application in a licensing context.

The PRIMUM® screens are the exclusive property of the National Board of Medical Examiners® ("NBME®"), Copyright © 1988-2008 by NBME. PRIMUM®, the United States Medical Licensing Examination® and USMLE® are federally registered trademarks of the NBME. The NBME is not responsible for the content of this article

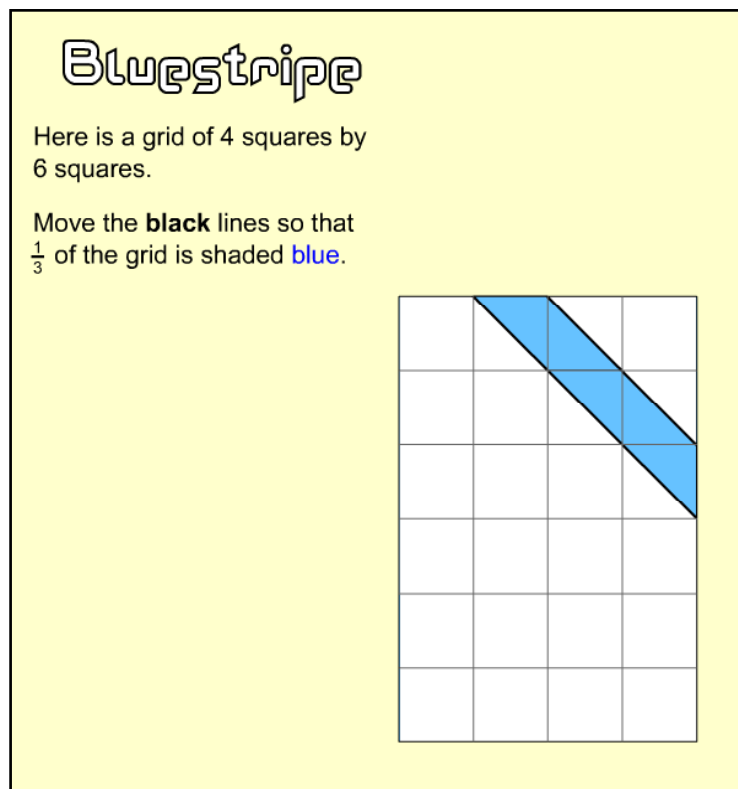
## World Class Tests

<b>Publisher:</b>	Qualifications and Curriculum Authority (QCA) and the UK Education Department, currently the Department for Children, Schools and Families (DCSF).
<b>Contact:</b>	QCA and/or Martin Ripley martin.ripley1@btinternet.com
<b>Brief details</b>	
<p>In 2000 England's Department for Education commissioned the development of new computer-based tests of problem-solving, in the domains of mathematics, science, design and technology. These tests were designed for world-wide application; they were designed to make creative use of computer technology. And they were designed to set new benchmarks in the design of assessments of students' thinking and ability to apply a range of techniques to solve novel and unexpected problems. These tests became known as World Class Tests and were adapted for children aged 8-14. These tests are now sold commercially under license in the Far East.</p>	
<b>The innovation</b>	
<p>The focus for innovation for World Class Tests was in the use of technology to design authentic, challenging and interesting problem solving tasks for students. In particular, the designers (Martin Ripley, Professor Hugh Burkhardt, Professor Jim Ridgway, Malcolm Swan, Sean McCusker, Peter Pool, John Threlfall, Jeremy Tafler and many others) sought to design:</p> <ul style="list-style-type: none"> <li>• Screens that were engaging, appealing and motivating to students;</li> <li>• Problems which included real-to-life contexts;</li> <li>• Problems which were complex to solve, requiring insight, perseverance and creativity.</li> </ul>	
<b>Description</b>	
<p>By definition, many transformative assessments will operate in non-traditional skills domains – such as problem solving, or team working, communication or innovation. One set of tests that has sought to create and develop a new domain are the World Class Tests developed in the UK. These designed were inspired by the UK government. The design brief was to create on-computer problem-solving assessments for highly able 8-13 year olds from around the world. The World Class Tests cover a number of problem-solving domains. Peter Pool led the development of mathematical problem-solving World Class Tests at the University of Leeds. He describes the brief for the tests. "The assessments are not about seeing how much mathematics has been covered - the questions do not require knowledge of</p>	

mathematical content beyond normal expectations for students at ages 9 and 13, so acceleration through the curriculum brings no great advantage. The questions are about how deeply the mathematics is understood and they offer success to those who can bring insight, perseverance and flexibility of thought to a question.” (Pool 2006)

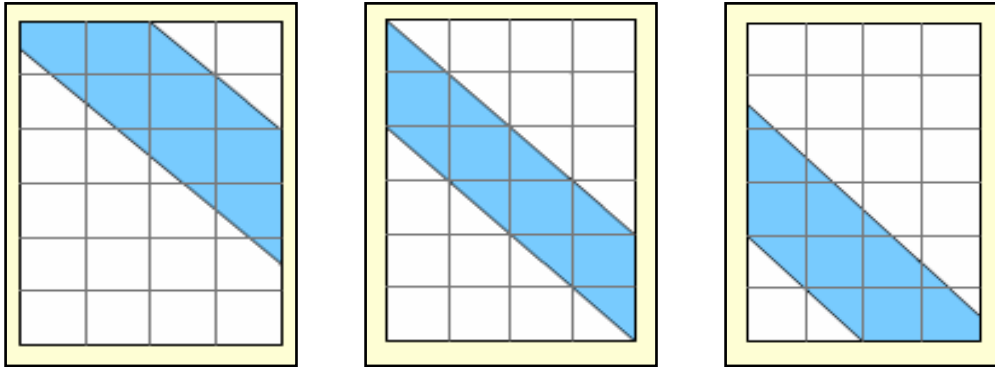
One example of a World Class Tests item is **Bluestripe**. The following description is taken from Peter Pool's paper. This is a grid of squares with an adjustable shaded band. Each of the two slant edges of the shaded area can be moved parallel to its starting position.

*Example of a World Class Tests Question - Bluestripe*



Moving either or both of the parallel sides can cause the shaded area to change its shape from trapezium to pentagon to hexagon to parallelogram. Its area needs to be 8 squares (from the information in the question), but it is difficult to operationalise this fact in any formulaic way over such a wide range of shapes, though whole and half squares are easily countable in individual cases. There are a number of approaches available - from counting squares to using the formula for the area of a parallelogram or trapezium, (knowing that the diagonal of a square is  $\sqrt{2}$  x side length). But none of these is likely to be deployed before some exploration has taken place using the interactivity of the diagram. This allows students to see the shapes that are possible, to recognise those that have areas that are easy to calculate, to get a sense of an approximate answer or to notice other aspects that might suggest a way forward. Theoretically, there is an infinite number of possible solutions

though most would require precise measurements that are not possible for the student on a computer screen - itself an additional factor for the student to take into account. Three of the more likely solutions are:



In the middle diagram the shaded band can be seen as having four identical vertical parallelogram sections, each one square wide and having a structure of one whole square and two half squares. In the left and right hand diagrams the small shaded part squares can be matched to small white sections to make complete shaded squares.

None of these solutions requires advanced understanding of how to calculate areas of shapes. What is more useful is the ability take advantage of the interactivity to recognise useful features that can be investigated and from which a strategy can be evolved. It is worth noting that it would be almost impossible to ask this question on paper in such a way that a student would be confident she had understood the procedure; the practicalities of then doing the question on paper raise further issues of manageability. In this sense, the mathematics here is 'new' in so far as it would not (or could not) be presented in a conventional paper assessed curriculum, though the question itself remains very accessible to anyone who understands conventional mathematics.

Research completed by Valsa Koshy and Ron Casey indicates clearly that World Class Tests do assess problem solving skills not traditionally assessed. (Koshy 2001) They found that students who have highly sophisticated and well-developed problem-solving strategies perform well on World Class Tests. However, they also found that the same students did not perform as well on traditional test of mathematics. Koshy and Casey developed the term "submerged talent" to capture the notion of World Class Tests capability to identify latent problem-solving talent at a very high level of refinement. Arguably young children showing the skills required for high performance on World Class Tests are already evidencing the types of skills required in the 21<sup>st</sup> century.

There remain a number of issues with World Class Tests which, nine years on from 2000 when they were first conceived, are now more susceptible of technology-led solutions. These issues include the following:

The tests do not assess elegance of student responses - any correct solution is accepted.

Each problem activity is large compared to more traditional assessment items, and each test has a small number of problem solving contexts. This has significant implications for the psychometric analysis of student performance - each problem takes a significant amount of testing time to complete.

As originally designed in 2000-02, many of the problem solving activities require students to respond on paper. Paper responses were used for aspects of problem solving which required:

- That students should be able to choose how best to express their understanding or to explain their solutions;
- Where the computer keyboard provided insufficient range for student responses - for example, where a drawing, a sketch, or formula might be required.

#### **Benefits**

World Class Tests have shown that:

- It is possible, affordably and reliably, to design micro-world, flash based test items which are engaging for young students, assessing their problem solving skills and capabilities.
- Performance in the tests is not strongly related to performance on other, more traditional and academic tests.
- Innovative tests can be designed which reflect positively significant aspects of learning that educators strive to find in the classroom in the 21<sup>st</sup> century.

#### **Finding out more**

[An information site, with further interactive examples of World Class Tests, is being developed and will be made available at.....TBC].

For more information, contact Martin Ripley at [martin.ripley1@btinternet.com](mailto:martin.ripley1@btinternet.com)



## Virtual Performance Assessment

Publisher:	Harvard University, Cambridge, MA, USA.
Contact:	<a href="mailto:jody_clarke@mail.harvard.edu">jody_clarke@mail.harvard.edu</a> ; <a href="mailto:chris_dede@harvard.edu">chris_dede@harvard.edu</a>
<b>Brief details</b>	
<p>The Virtual Performance Assessment (VPA) project is a three year grant (9/2008-9/2011) funded by the Institute of Educational Sciences (IES) to study the feasibility of using immersive technologies to deliver virtual performance assessments that measure science inquiry knowledge and skills, as defined by the National Science Education Standards (NSES). We are developing the assessments for use in school settings as a standardized component of an accountability program. Our goal is to develop three assessments in the context of life science that appear different on the surface, but are all measuring the same inquiry process skills. Each assessment will take place in a different type of ecosystem, and students will investigate authentic ecological problems as they engage in the inquiry process.</p> <p>Performance of scientific inquiry is a primary objective for middle school science, as stated by the National Science Education Standards (NSES). Research documents that paper-and-pencil tests, such as the US National Assessment of Educational Progress (NAEP), do not effectively measure inquiry skills (Quellmalz, Kreikemeier, Haydel-DeBarger, &amp; Haertel, 2007). In the 1980s and 1990s, researchers tried to develop hands-on, physical performance assessments of scientific inquiry, as a more authentic approach to assessing inquiry. However, the hands-on assessments encountered numerous issues, including technical, resource, and reliability problems.</p> <p>We believe that advances in technology, cognition, and mathematical modeling offer potential for assessing students' performance of scientific inquiry with an Immersive Virtual Environment (IVE). Research over the last decade on pedagogical applications of IVEs (Dede, 2009) has documented increased student engagement in authentic inquiry tasks (problem finding and experimental design) and increased student self-efficacy (Ketelhut, 2007; Nelson, 2007; Clarke &amp; Dede, 2009). Thus, we are building off prior research on assessing science inquiry via IVEs, as well as using Evidence Centered Design (Mislevy, Steinberg, &amp; Almond, 2003) to create an evidentiary argument for making inferences about students' performance. We believe that virtual performance assessments could mitigate limitations encountered historically with both physical performance assessments and paper and pencil item-based tests.</p>	
<b>The innovation</b>	
<p>The Virtual Performance Assessment project utilizes innovations in technology and assessment to address the problem of measuring a student's ability to perform scientific inquiry to solve a problem. Three innovative aspects of the VPA are discussed below:</p>	

<ul style="list-style-type: none"> <li>• IVE - The VPA uses an IVE to provide each student with a 3-D virtual environment for performing scientific inquiry to solve an authentic problem. The student takes on the role of a scientist and collects data that can help solve an ecological problem. The VPA assesses a student's ability to formulate an argument based on data s/he collects. The student's arguments are expressed in multiple representations through various digital media, including concept maps for cause and effect chains, data collection tools for gathering evidence, and conclusion tools for submitting hypotheses and causal statements, then selecting the most salient data justifying these.</li> <li>• Logfiles - The VPA makes a student's thinking visible by capturing his or her actions in the virtual environment. Everything the student does is captured and stored in logfiles. We will use this data to find patterns for the most and least effective strategies students used to gather data and solve the problem. This data will also allow us to capture inquiry trajectories.</li> <li>• Evidence Centered Design (ECD) - ECD is being used as a framework for being able to make valid inferences about a student's scientific inquiry skills based on their actions within the virtual environment. The ECD process provides a framework for statistically linking the desired knowledge, skills, and abilities with the task details and with the student model variables.</li> </ul>
<p><b>Description</b></p> <p>VPA offers authentic ecological problems for a student to investigate. The student can walk around to make observations and inferences, sorting out the most salient parts of a complex situation. After collecting data, the student then must distinguish the most important data values that can help solve the problem. S/he can create concept maps, rank data by importance, make predictions, and experiment with the virtual environment by altering it (e.g., removing a glacier), then collecting data post-change and submit a conclusion about whether altering the environment resulted in an improvement in the ecosystem. We will assess each student on his or her ability to interpret data that individual collects, as well as the data of simulated researchers. Further, we will assess each student on both product and process. Product assessment includes the student's scores on 5 tasks, each including a number of various sub-tasks. Process assessment consists of measuring the student's ability to perform scientific inquiry to solve the problem. This measure is an example of the proposed principle for 21st Century Standards and Assessments that states assessments should "be largely performance-based."</p>
<p><b>Benefits</b></p>

VPA is consistent with numerous principles listed in the 21<sup>st</sup> Century Standards and Assessments:

- *Ability to formulate one's arguments, in speaking or writing, in a convincing manner and take full account of other viewpoints, whether expressed in written or oral form.*

VPA assesses a student's ability to formulate an argument based on data that the student collected. The student's argument is expressed in multiple representations through the use of various digital media, including concept maps for cause and effect chains, data collection tools for gathering evidence, and conclusion tools for submitting hypotheses and causal statements, then selecting the most salient data that support these.

- *Skills needed to use aids (such as notes, schemes, maps) to produce, present or understand complex texts in written or oral form (speeches, conversations, instructions, interviews, debates).*

To score well in the VPA assessment, a student must possess critical thinking and scientific inquiry skills. In addition, s/he will need to be systematic in data collection strategy, due to having limited time and an abundance of potential data to collect. In addition, the student will need to interpret graphs and visual cause/effect chains. To level the playing field in students' technology skills, the VPA assessment is preceded by a sandbox activity, which will set the context for the assessment and provide a student with brief training on how to use the VPA data collection tools, navigate, and request help.

- *Prioritize, plan and manage work to achieve the intended result*

The student will have limited time and therefore must strategically manage data collection efforts.

- *Ability to search, collect and process (create, organize, distinguish relevant from irrelevant, subjective from objective, real from virtual) electronic information, data and concepts and to use them in a systematic way;*

The student searches for specific data relevant to solving an ecological problem, then must identify what data is crucial to solving the problem.

- *Ability to use appropriate aids presentations, graphs, charts, maps) to produce, present or understand complex information;*

The VPA requires the student to create conclusions (hypotheses, causal models, etc.) using performance palettes, such as concept maps, fill-in the blank sentences,

<p>etc.</p> <ul style="list-style-type: none"> <li>• <i>Ability to use information to support critical thinking, creativity and innovation in different contexts at home, leisure and work.</i></li> </ul> <p>The student must use critical thinking skills and be strategic in data collect to reach and justify the conclusions submitted.</p> <ul style="list-style-type: none"> <li>• <i>Ability to search, collect and process written information, data and concepts in order to use them in study and to organize knowledge in a systematic way. Ability to distinguish, in listening, speaking, reading and writing, relevant from irrelevant information.</i></li> </ul> <p>In the VPA, a student immersively explores and collects data in a virtual world with an ecological problem. The student presents hypotheses, causal models, and reflections using various digital tools. In particular, the student must select the most salient data to use to supporting their conclusions.</p>
<b>Finding out more</b>
<p>Visit the project at: <a href="http://virtualassessment.org/index.html">http://virtualassessment.org/index.html</a>.</p>

## Citations

- Clarke, J., & Dede, C. (2009). Design for Scalability: A Case Study of the River City Curriculum. *Journal of Science Education and Technology*.
- Dede, C. (2009). Immersive interfaces for engagement and learning. *Science*, 323(5910), 66-69.
- Ketelhut, D. J. (2007). The impact of student self-efficacy on scientific inquiry skills: An exploratory investigation in River City, a multi-user virtual environment. *The Journal of Science Education and Technology*, 16 (1), 99–111.
- Mislevy, R.J., Steinberg, L.S., & Almond, R.G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, 1, 3-62.
- Quellmalz, E., Kreikemeier, P., DeBarger, A. H., & Haertel, G. (2007). A study of the alignment of the NAEP, TIMSS, and New Standards Science Assessments with the inquiry abilities in the National Science Education Standards. Presented at AERA 2007.

## References and further reading

- Ainley, J., Fraillon, J., Freeman, C., & Mendelovits, J. (2006, April). *Assessing information and communication technology literacy in schools*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.
- Anderson, R. (2009, April). "A Plea for '21st Century Skills' White Paper to Include Social and Civic Values." Memorandum to Assessment and Teaching of 21st Century Skills Conference, San Diego, CA.
- Bennett, R.E. (2002). Inexorable and inevitable: The continuing story of technology and assessment. *Journal of Technology, Learning, and Assessment*, 1(1).
- Bennett, R. E. (2008). *Technology for large-scale assessment*. Educational Testing Service Research Memorandum. Princeton, NJ.
- Bennett, R. E., & Gitomer, D. H. (2009). Transforming K-12 assessment. In C. Wyatt-Smith & J. Cumming (Eds.), *Assessment issues of the 21st Century*. New York, NY: Springer Publishing Company.
- Bennett, R. E., Jenkins, F., Persky, H., & Weiss, A. (2003). Assessing complex problem solving performances. *Assessment in Education: Principles, Policy & Practice*, 10, 347-360.
- Black, B. (2008, September). Critical thinking – a definition and taxonomy for Cambridge Assessment: Supporting validity arguments about critical thinking assessments administered by Cambridge Assessment. Paper presented at 34<sup>th</sup> International Association of Educational Assessment Annual Conference, Cambridge, England.
- Center for Assessment and Evaluation of Student Learning (CAESL). (2008). "Model of a comprehensive assessment system." Oakland, CA.
- Center for Research on Educational Testing. (2009). *Generic skills in Japan*. Tokyo, Japan.
- Clift, S. (2002). *21st Literacy Summit White Paper*. Retrieved from [www.mail-archive.com/dowire@tc.umn.edu/msg00434.html](http://www.mail-archive.com/dowire@tc.umn.edu/msg00434.html)
- Collegiate Assessment of Academic Proficiency (CAPP). (2009). *Critical thinking sample passages and items*. Retrieved from <http://www.act.org/caap/sample/thinking.html>
- Deakin Crick, R. D., Broadfoot, P., & Claxton, G. (2004). Developing an effective lifelong learning inventory: The ELLI project. *Assessment in Education: Principles, Policy & Practice*, 11, 247-318.
- Dede, C. (Forthcoming). Technological supports for acquiring 21st century skills. Paper commissioned for the International Encyclopedia of Education.
- Erstad, O. (2006). A new direction? Digital literacy, student participation and curriculum reform in Norway. *Education and Information Technologies*, 11(3-4), 415-429.
- Fadel, C., Honey, M., & Pasnik, S. (2007, May 18). Assessment in the age of innovation, *Education Week*. Retrieved from <http://www.edweek.org/ew/articles/2007/05/23/38fadel.h26.html>
- Harlen, W., & R. Deakin Crick, R. (2003). Testing and motivation for learning. *Assessment in Education: Principles, Policy & Practice*, 10, 169-208.
- Honey, M., Pasnik, S., & Fasca, C. (2006, Spring). Is having the right answer enough? Developing new forms of assessment for 21st century learning. *Threshold*, 22-26.
- Hoskins, B., & Fredriksson U. (2008). *Learning to learn: What is it and can it be measured?* European Union. Joint Research Centre, Italy. Retrieved from [http://crell.jrc.ec.europa.eu/L2L/learning%20to%20learn%20what%20is%20it%20and%20can%20it%20be%20measured\\_ver5.pdf](http://crell.jrc.ec.europa.eu/L2L/learning%20to%20learn%20what%20is%20it%20and%20can%20it%20be%20measured_ver5.pdf)

- International ICT Literacy Panel. (2002). *Digital transformation: A framework for ICT literacy*. Princeton, NJ: Educational Testing Service.
- Jewitt, C. (2003). Re-thinking assessment: Multimodality, literacy and computer-mediated learning. *Assessment in Education: Principles, Policy & Practice*, 10, 83-102.
- Kozma, R. B. (Ed.) (2003). *Technology, innovation, and educational change: A global perspective*. Eugene, OR: International Society for the Evaluation of Educational Achievement.
- Kulik, J.A. (1994). Meta-analytic studies of findings on computer-based instruction. In E. L. Baker & H. F. O'Neil (Eds.), *Technology assessment in education and training*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Lesgold, A. (2009). *Better schools for the 21st century: What is needed and what will it take to get improvement*. University of Pittsburgh. Retrieved from <http://www.learningpolicycenter.org/data/publications/Alan-Lesgold-Better-Schools-for-the-21st-Century.pdf>
- Lin, S. S. J., Liu, E. Z. F., & Yuan, S. M. (2001). Web-based peer assessment: Feedback for students with various thinking styles. *Journal of Computer Assisted Learning*, 17, 420-432.
- Loader, B. (Ed.). (2007). *Young citizens in the digital age: Political engagement, young people and new media*. London: Routledge.
- McFarlane, A. (2001). Perspectives on the relationships between ICT and assessment. *Journal of Computer Assisted Learning*, 17, 227-234.
- McFarlane, A. (2003). Assessment for the digital age. *Assessment in Education: Principles, Policy & Practice*, 10, 261-266.
- Nunes, C. A. A., Nunes, M. M. R., & Davis, C. (2003). Assessing the inaccessible: Metacognition and attitudes. *Assessment in Education: Principles, Policy & Practice*, 10, 375-388.
- O'Neil, H. F., Chuang, S., & Chung, G. K. W. K. (2003). Issues in the computer-based assessment of collaborative problem solving. *Assessment in Education: Principles, Policy & Practice*, 10, 361-374.
- Oxford Cambridge and RSA Examinations. (2008). *GCE critical thinking*. Cambridge, England.
- Pellegrio, J. W., Chudowsky, N. & Glaser, R., editors (2001). *Knowing What Students Know*. Washington, DC: National Academy Press.
- Poggio, J., Glasnapp, D. R., Yang, X., & Poggio, A. J. (2005). A comparative evaluation of score results from computerized and paper and pencil mathematics testing in a large scale state assessment program. *Journal of Technology, Learning, and Assessment*, 3(6).
- Pommerich, M. (2004). Developing computerized versions of paper-and-pencil tests: Mode effects for passage-based tests, *Journal of Technology, Learning and Assessment*, 2(6).
- Quellmalz, E. S., & Kozma, R. (2003). Designing assessments of learning with technology. *Assessment in Education: Principles, Policy & Practice*, 10, 389-408.
- Raikes, N., & Harding, R. (2003). The horseless carriage stage: Replacing conventional measures. *Assessment in Education: Principles, Policy & Practice*, 10, 267-278.
- Ridgway, J., & McCusker, S. (2003). Using computers to assess new educational goals. *Assessment in Education: Principles, Policy and Practice*, 10(3), 309-328.
- Ridgway, J., McCusker, S., & Pead, D. (2004). *Literature review of e-assessment (report 10)*. Bristol, England: Futurelab.
- Russell, M. (1999). Testing on computers: A follow-up study comparing performance on computer and on paper. *Education Policy Analysis Archives*, 7(20). Retrieved from <http://epaa.asu.edu/epaa/v7n20>

- Russell, M., Goldberg, A., & O'Connor, K. (2003). Computer-based testing and validity: A look into the future. *Assessment in Education: Principles, Policy & Practice*, 10, 279-294.
- Russell, M., & Haney, W. (2000). Bridging the gap between testing and technology in schools. *Education Policy Analysis Archives*, 8(19). Retrieved from <http://epaa.asu.edu/epaa/v8n19.html>
- Sefton-Green, J., & Sinker, R. (Eds.). (2000). *Evaluating creativity: Making and learning by young people*. London: Routledge.
- Shute, V., Masduki, M., Donmez, O., Dennen, V. P., Kim, Y-J., Jeong, A. C., & Wang, C-Y. (Forthcoming). Modeling, assessing and supporting key skills within game environments. To appear in D. Ifenthaler, P. Pirnay-Dummer, N. M. Seel (Eds.), *Computer-based diagnostics and systematic analysis of knowledge*. New York, NY: Springer-Verlag.
- Shute, V. J., Ventura, M., Bauer, M., & Zapata-Rivera, D. (2009.) Melding the power of serious games and embedded assessment to monitor and foster learning: Flow and grow. New York: Routledge.
- Somekh, B., & Mavers, D. (2003). Mapping learning potential: Students' conceptions of ICT in their world. *Assessment in Education: Principles, Policy & Practice*, 10, 409-420.
- Voluntary System of Accountability (VSA) Program. (2009). *Learning outcomes workshops*. Retrieved from <http://www.voluntarysystem.org>
- Walton, S. (2005). *The eVIVA project: Using e-portfolios in the classroom*. BETT. Retrieved 6/7/07 from [www.qca.org.uk/downloads/10359\\_eviva\\_bett\\_2005.pdf](http://www.qca.org.uk/downloads/10359_eviva_bett_2005.pdf)
- Wasson, B., Ludvigsen, S., & Hoppe, U. (Eds.). (2003). Designing for change in networked learning environments: Proceedings of the International Conference on Computer Support for Collaborative Learning 2003. *Computer-Supported Collaborative Learning Series: Vol 2*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Woodward, H., & Nanlohy, P. (2004). Digital portfolios in pre-service teacher education. *Assessment in Education: Principles, Policy & Practice*, 11, 167-178.