Changing Assessment — Towards a New Assessment Paradigm Using ICT

Christine Redecker & Øystein Johannessen

Rethinking 21st Century Assessment

We are currently living in an era of accelerating change as concerns not only technological developments, but also society as a whole. Hence, the skills and competences needed for work and life in the 21st century are continuously evolving. Policy is reacting to these changes by calling for education to focus on the development of Key Competences for Lifelong Learning (Council of the European Union, 2006). Moreover, the recent 'Rethinking Education Strategy' (http://ec.europa.eu/education/news/20121120_en.htm) highlights the need to focus on the development of transversal and basic skills at all levels, especially entrepreneurial and IT skills. However, learning processes and goals can only change if assessment also changes. Assessment is an essential component of learning and teaching, as it allows the quality of both teaching and learning to be judged and improved (Ferrari, Cachia, & Punie, 2009). It often determines the priorities of education (NACCCE, 1999), it always influences practices and affects learning (Ellis & Barrs, 2008). Changes in curricula and learning objectives are ineffective if assessment practices remain the same (Cachia *et al.*, 2010), as learning and teaching tend to be modelled against the test (NACCCE, 1999).

Assessment is usually understood to have a formative and a summative purpose. Formative assessment aims to gather evidence about pupils' proficiency in order to influence teaching methods and priorities, whereas summative assessment is used to judge pupils' achievements at the end of a programme of work (NACCCE, 1999). For it to be effective, students must become self-regulated learners (Nicol & MacFarlane-Dick, 2006) who monitor the quality of their work, a capacity which must be fostered by the learning environment (Sadler, 1989). The role of assessment in facilitating good learning environments was highlighted by the OECD Innovative Learning Environments Project (OECD, 2010). Summing up the transversal characteristics of learning environments:

> Formative assessment is a central feature of the learning environment of the 21st century. Learners need substantial, regular and meaningful feedback; teachers need it in order to understand who is learning and how to orchestrate the learning process.

Assessment procedures in formal education and training have traditionally focused on examining knowledge and facts through formal testing (Cachia *et al.*, 2010) and do not easily lend themselves to grasping 'soft skills'. Lately, however, there has been a growing awareness that curricula — and with them assessment strategies — need to be revised to more adequately reflect the skills needed for life in the 21st century. The evolution of information and communication technologies (ICT) is deeply re-shaping society, giving rise to new competence needs. Skills such as problem-solving, reflection, creativity, critical thinking, learning to

learn, risk-taking, collaboration, and entrepreneurship are becoming increasingly important (Redecker *et al.*, 2010). The relevance of these '21st century skills' (Binkley *et al.*, 2012) is recognised in the European Recommendation on Key Competences for Lifelong Learning (2006) which emphasises their transversal and over-arching role. To foster and develop these skills, assessment strategies should go beyond testing factual knowledge and capture the less tangible themes underlying all Key Competences. At the same time, assessment strategies need to be better harmonised with 21st century learning approaches by re-focusing on the importance of providing timely and meaningful feedback to both learners and teachers.

This article discusses how Information and Communication Technologies can support this shift towards 21st century assessment strategies and what needs to be done to ensure that technological advances support and foster pedagogical innovation (See Istance & Kools in this issue pp. 43–57, and Bocconi *et al*, pp. 113–130). Based on an extensive review of the literature, it provides an overview of current ICT-enabled assessment practices, with a particular focus on the more recent developments of ICT-enhanced assessment tools that allow the recognition of 21st century skills. The article also refers to relevant cases of eAssessment, looks into examples of the potential of emerging technologies for eAssessment and discusses some relevant innovation and policy issues for eAssessment.

Looking Beyond the e-Assessment Era

Currently, the world of education is influenced by a plethora of emerging technologies (http://c4lpt.co.uk/top-100-tools-2012/; www.edtechmagazine.com/k12/article/2012/11/6-hot-trends-educational-technology-infographic). Various studies show that the number of available technologies is very high and have reached a varying degree of maturity (Coalition, 2012; Gartner Research, 2012). Many have similar lifecycles spanning embryonic stages, early adoption and great anticipations, mainstreaming and end-of-lifecycle. Few originally come from the field of education. They are often imported from either business or consumer electronics and then adapted to the needs of educational practice.

As the Horizon Project (Johnson, Adams, & Cummins, 2012a, 2012b; Johnson *et al*, 2011) has documented, the landscape of emerging technologies for teaching and learning is changing. Some changes are relatively incremental, others are more disruptive. Some of the emerging technologies that we either have observed in educational practice or which will enter education in the next few years have promising potential for assessment. At present, we stand at the crossroads of two 'assessment paradigms' and lack a pedagogical vision of how to move from the old one, the era of computer-based testing, to the new one, the era of embedded assessment.

Technology-enabled assessment develops in various stages. This development bears some resemblance to the SAMR-model (Substitution — Augmentation — Modification — Redefinition) which was coined by Ruben Puentedura and describes the different stages of technology adoption (and possible enhancement) in a domain of practice, e.g. teaching and learning. Figure 1 shows the various stages of the technology adoption lifecycle:

Similarly, at the end of the 1980s, Bunderson, Inouye and Olsen (1989) forecast four generations of computerised educational measurement, namely:
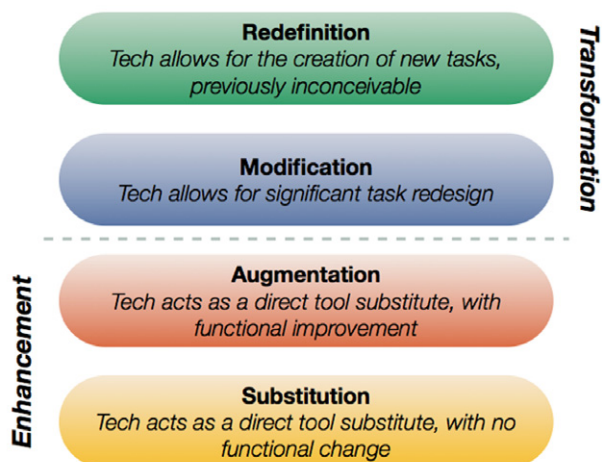
FIGURE 1. SAMR-Model (Puentedura, 2012)

Generation 1: Computerised testing (administering conventional tests by computer);

Generation 2: Computerised adaptive testing (tailoring the difficulty or contents or an aspect of the timing on the basis of examinees' responses);

Generation 3: Continuous measurement (using calibrated measures to continuously and unobtrusively estimate dynamic changes in the student's achievement trajectory);

Generation 4: Intelligent measurement (producing intelligent scoring, interpretation of individual profiles, and advice to learners and teachers by means of knowledge bases and inferencing procedures).

Interestingly, these predictions are not far off the mark. The first two generations of eAssessment or Computer-Based Assessment (CBA), which should more precisely be referred to as Computer-Based *Testing*, have now become mainstream. The main challenge now lies in making the transition to the latter two, the era of Embedded Assessment, which is based on the notion of 'Learning Analytics', i.e. the interpretation of data about students' proficiency in order to assess academic progress, predict future performance, and tailor education to individual students. Although Learning Analytics are still at an experimental and development stage, embedded assessment could become a reality within the next five years (Johnson *et al.*, 2011).

However, the transition from computer-based testing to embedded assessment, from the phase of enhancement to the era of transformation, requires technological advances to be complemented with a conceptual shift in assessment paradigms. While the first two generations of CBA centre on the notion of *testing* and the use of computers to improve the efficiency of testing procedures, generations 3 and 4 seamlessly integrate holistic and personalised assessment into *learning*. Embedded assessment allows learners to be continuously monitored and guided by the electronic environment which they use for their learning activities, thus merging formative and summative assessment within the learning process. Ultimately, with generation 4, learning systems will be able to provide instant and valid feedback and advice to learners and teachers concerning future learning strategies, based on
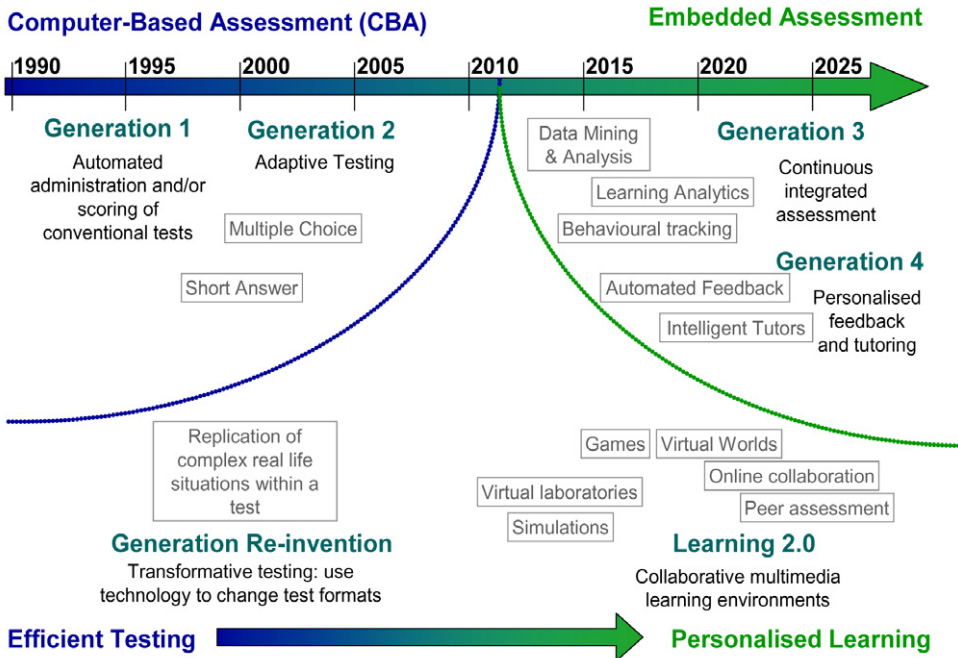
FIGURE 2. Current and future e-Assessment strategies
Source: IPTS on the basis of Bennett, 2010; Bunderson, et al., 1989; Martin, 2008).

the learners' individual learning needs and preferences. Explicit testing could thus become obsolete.

This conceptual shift in the area of eAssessment is paralleled by the overall pedagogical shift from knowledge- to competence-based learning (Eurydice, 2012) and the recent focus on transversal and generic skills, which are less susceptible to generation 1 and 2 e-Assessment strategies. Generation 3 and 4 assessment formats may offer a viable avenue to capture the more complex and transversal skills and competences that are crucial for work and life in the 21st century. However, to seize these opportunities, assessment paradigms need to become enablers of more personalised and targeted learning processes. Hence, the question is: how we can make this shift happen?

What we currently observe is that these two, in our opinion, conceptually different eAssessment approaches — the 'Explicit Testing Paradigm' and the 'Embedded Assessment Paradigm' — develop in parallel to accommodate more complex and authentic assessment tasks that better reflect 21st century skills and more adequately support the recent shift towards competence-based curricula.

The Explicit Testing Paradigm

Mainstream eAssessment Strategies

First and second generation tests have led to a more effective and efficient delivery of traditional assessments (Martin, 2008). More recently, assessment tools have been enriched to include more authentic tasks and allow for the assessment of

constructs that have either been difficult to assess or have emerged as part of the information age (Pellegrino, 2010). As the measurement accuracy of all these test approaches depends on the quality of the items it includes, item selection proce-dures — such as the Item Response Theory or mathematical programming — play a central role in the assessment process (El-Alfy & Abdel-Aal, 2008). First gen-eration computer-based tests are already being administered widely for a variety of educational purposes, especially in the US (Csapó *et al*, 2010), but increasingly also in Europe (Moe, 2009). Generation adaptive tests select test items based on the candidates' previous response, allow for a more efficient administration mode (less items and less testing time), while keeping measurement precision (Martin, 2008). Different algorithms have been developed for the selection of test items. The most well-known is CAT, which is a test where the algorithm is designed to provide an accurate point estimation of individual achievement (Thompson & Weiss, 2009). CAT tests are very widespread, in particular in the US, where they are used for assessment at primary and secondary school level (Bennett, 2010; Bridgeman, 2009; Csapó *et al*., 2010); and for admission to higher education (Bridgeman, 2009). Adaptive tests are also used in European countries, for instance the Netherlands (Eggen & Straetmans, 2009) and Denmark (Wandall, 2009).

Reliability of eAssessment

Advantages of computer-based testing over traditional assessment formats include: paperless test distribution and data collection; efficiency gains; rapid feedback; machine-scorable responses; and standardised tools for examinees such as calcu-lators and dictionaries (Bridgeman, 2009). Furthermore computer-based tests tend to have a positive effect on students' motivation, concentration and perform-ance (Garrett *et al*, 2009). More recently, they have also provided learners and teachers with detailed reports that describe strengths and weaknesses, thus sup-porting formative assessment (Ripley, 2009a).

However, the reliability and validity of scores have been major concerns, particularly in the early phases of eAssessment, given the prevalence of multiple-choice formats in computer-based tests. Recent research indicates that scores are generally higher in multiple choice tests than in short answer formats (Park, 2010). Some studies found no significant differences between student performance on paper and on screen (Hardré *et al*.,2007; Ripley, 2009a), whereas others indicate that paper-based and computer-based tests do not necessarily measure the same skills (Bennett, 2010; Horkay *et al*., 2006).

Current research concentrates on improving the reliability and validity of test scores by improving selection procedures for large item banks (El-Alfy & Abdel-Aal, 2008), increasing measurement efficiency (Frey & Seitz, 2009), taking into account multiple basic abilities simultaneously (Hartig & Höhler, 2009), and developing algorithms for automated language analysis that allow the electronic scoring of long free text answers.

Automated Scoring

One of the drivers of progress in eAssessment has been the improvement of automatic scoring techniques for free text answers (Noorbehbahani & Kardan, 2011) and written text assignments (He, Hui, & Quan, 2009). Automated scoring could dramatically reduce the time and costs of the assessment of

complex skills such as writing, but its use must be validated against a variety of criteria for it to be accepted by test users and stakeholders (Weigle, 2010). Assignments in programming languages or other formal notations can already be automatically assessed (Amelung, Krieger, & Rösner, 2011). For short-answer free-text responses of around one sentence, automatic scoring has also been shown to be at least as good as human markers (Butcher & Jordan, 2010). Similarly, automated scoring for highly predictable speech, such as a one sentence answer to a simple question, correlates very highly with human ratings of speech quality, although this is not the case with longer and more open-ended responses (Bridgeman, 2009). Automated scoring is also used for scoring essay-length responses (Bennett, 2010) where it is found to closely mimic the results of human scoring[1]: the agreement of an electronic score with a human score is typically as high as the agreement between two humans, and sometimes even higher (Bennett, 2010; Bridgeman, 2009; Weigle, 2010). However, these programmes tend to omit features that cannot be easily computed, such as content, organisation and development (Ben-Simon & Bennett, 2007). Thus, while there is generally a high correlation between human and machine marking, discrepancies are greater for essays of more abstract qualities (Hutchison, 2007). A further research line aims to develop programmes which mark short-answer free-text and give tailored feedback on incorrect and incomplete responses, inviting examinees to repeat the task immediately (Jordan & Mitchell, 2009).

Transformative Testing

The transformational approach to Computer-Based Assessment uses complex simulation, sampling of student performance repeatedly over time, integration of assessment with instruction, and the measurement of new skills in more sophisticated ways (Bennett, 2010). Here, according to Ripley (2009a), the test developer redefines assessment and testing approaches in order to allow for the assessment of 21st century skills. By allowing for more complex cognitive strategies to be assessed while remaining grounded in the testing paradigm, these innovative testing formats could facilitate the paradigm shift between the first two and the last two generations of eAssessment, e.g. between explicit and implicit assessment.

A good example of transformative tests is computer-based tests that measure process by tracking students' activities on the computer while answering a question or performing a task, such as in the ETS iSkills test. However, experience from this and other trials indicates that developing these tests is far from trivial (Lent, 2008). One of the greatest challenges for the developers of transformative assessments is to design new, robust, comprehensible and publicly acceptable means of scoring students' work (Ripley, 2009a).

In the US, a new series of national tests is being developed which will be implemented in 2014–15 for maths and English language arts and take up many of the principles of transformative testing (ETS, 2012).

CBA for Formative Assessment

Computer-based assessment can also be used to support formative and diagnostic testing. The Norwegian Centre for ICT in Education (https://iktsenteret.no/english/) has been piloting a diagnostic test for digital literacy in Oslo and Bergen,

the two largest cities in Norway, with 30 schools with 800 students participating in 2012. It is based on a national framework which defines four 'digital areas': to acquire and process digital information, produce and process digital information, critical thinking and digital communication. Test questions are developed and implemented on different platforms with different functionalities and include simulation and interactive test items.

In Hungary, a networked platform for diagnostic assessment (www.edu.u-szeged.hu/~csapo/irodalom/DIA/Diagnostic_Asessment_Project.pdf) is being developed from an online assessment system developed by the Centre for Research on Learning and Instruction at the University of Szeged. The goal is to lay the foundation for a nationwide diagnostic assessment system for the grades 1 through 6. The project will develop an item bank in nine dimensions (reading, mathematics and science in three domains) as well as a number of other minor domains. According to the project management, it has a high degree of transferability and can be regarded as close to the framework in the PISA assessment.

The Embedded Assessment Paradigm

Unlike the Explicit Testing paradigm, the Embedded Assessment paradigm does away with tests and instead, via Learning Analytics, uses the data produced during the learning process as a basis for providing feedback and guidance to learners, teachers and parents. The 2011 edition of the Horizon report K-12 identified Learning Analytics as an emerging technology that was likely to enter mainstream practice in education on a 4–5 year horizon. However, at present, Learning Analytics is still in its infancy and is widely debated. An elaboration and mainstreaming of this technology will require a combination of data from various sources, and it raises some issues with regard to learner privacy and security (Johnson *et al*, 2011). On the other hand, its power lies in its potential to provide learners, teachers, tutors and parents with real-time information about the needs and progress of each individual learner which could enable rapid intervention and greater personalisation. The feasibility of Learning Analytics requires further analysis in a policy environment in which, on a global scale, both traditional and innovative approaches to assessment coexist. If prudently used, it could take formative eAssessment to a higher level.

Learning Analytics is already implemented in some technology-enhanced learning (TEL) environments where data-mining techniques can be used for formative assessment and individual tutoring. Similarly, virtual worlds, games, simulations and virtual laboratories allow the tracking of individual learners' activity and can make learning behaviour assessable. Assessment packages for Learning Management Systems are currently being developed to integrate self-assessment, peer-assessment and summative assessment, based on the automatic analysis of learner data (Florián *et al.*, 2010). Furthermore, data on student engagement in these environments can be used for embedded assessment, which refers to students engaging in learning activities while an assessment system draws conclusions based on their tasks (Ridgway & McCusker, 2008). Data-mining techniques are already used to evaluate university students' activity patterns in Virtual Learning Environments for diagnostic purposes. Analytical data mining can, for example, identify students who are at risk of dropping out or underperforming,[2] generate diagnostic and performance reports (www.socrato.com), assess

interaction patterns between students on collaborative tasks (http://research.uow.edu.au/learningnetworks/seeing/snapp/index.html), and visualise collaborative knowledge work (http://emergingmediainitiative.com/project/learning-analytics/). It is expected that, in five years' time, advances in data mining will enable one to interpret data concerning students' engagement, performance, and progress in order to assess academic progress, predict future performance, and revise curricula and teaching strategies (Johnson *et al.*, 2011).

Although many of these programmes and environments are still experimental in scope and implementation, a number of promising technologies and related assessment strategies could soon give rise to integrated assessment formats that comprehensively capture 21st century skills.

Automated Feedback

Research indicates that the closer the feedback to the actual performance, the more powerful its impact on subsequent performance and learner motivation (Nunan, 2010). Timing is the obvious advantage of Intelligent Tutoring Systems (ITSs) (Ljungdahl & Prescott, 2009) that adapt the level of difficulty of the tasks to the individual learners' progress and needs. Most programmes provide qualitative information on why particular responses are incorrect (Nunan, 2010). Although in some cases fairly generic, some programmes search for patterns in student work to adjust the level of difficulty in subsequent exercises according to needs (Looney, 2010). Huang *et al.* (2011) developed an intelligent argumentation assessment system for elementary school pupils which analyses the structure of students' scientific arguments posted on a moodle discussion board and issues feedback in case of bias. In a first trial, it was shown to be effective in classifying and improving students' argumentation levels and assisting them in learning the core concepts.

ITSs are being widely used in the US, where the most popular system 'Cognitive Tutor' provides differentiated instruction in mathematics which encourages problem-solving behaviour among half a million students in middle and high schools (Ritter *et al.*, 2010). It selects problems for each student at an adapted level of difficulty. Correct solution strategies are annotated with hints (http://carnegielearning.com/static/web_docs/2010_Cognitive_Tutor_Effectiveness.pdf). Research commissioned by the programme developers indicates that students who used Cognitive Tutor greatly outscored their peers in national exams, an effect that was especially noticeable in students with limited English proficiency or special learning needs (Ritter *et al.*, 2007). Research on the implementation of a web-based intelligent tutoring system 'eFit' for mathematics at lower secondary schools in Germany confirms this: children who used it significantly improved their arithmetic performance over a period of 9 months (Graff, Mayer, & Lebens, 2008).

ITSs are also used to support reading. SuccessMaker's Reader's Workshop (www.successmaker.com/Courses/c_awc_rw.html). and Accelerated Reader (www.renlearn.com/ar/) are two very popular commercial reading software products for primary education in the US. They provide ICT-based instruction with animations and game-like scenarios. Assessment is embedded and feedback is automatic and instant. Learning can be customised for three different profiles and each lesson can be adapted to students' strengths and weaknesses. These programmes have been evaluated as having positive impacts on learning (Looney, 2010).

Immersive Environments, Virtual Worlds, Games and Simulations

Immersive environments and games are specifically suitable for acquiring 21st century skills such as problem-solving, collaboration and inquiry because they are based on the fact that what needs to be acquired is not explicit but must be inferred from the situation (de Jong, 2010). In these environments, the learning context is similar to the contexts in which students will apply their learning, thus promoting inquiry skills; making learning activities more motivating; and increasing the likelihood that acquired skills will transfer to real-world situations (Means & Rochelle, 2010). It has been recognised that immersive game-based learning environments lead to significantly better learning results than traditional learning approaches (Barab *et al.*, 2009).

Assessment can be integrated in the learning process in virtual environments. In science education, for instance, computer simulations, scientific games and virtual laboratories provide opportunities for students to develop and apply skills and knowledge in more realistic contexts and provide feedback in real time. Dynamic websites, such as Web of Inquiry (www.webofinquiry.org), allow students to carry out scientific inquiry projects to develop and test their theories; learn scientific language, tools, and investigation practices; engage in self-assessment; and provide feedback to peers (Herrenkohl, Tasker, & White, 2011). Simulations provided by Molecular Workbench (http//mw.concord.org/modeler/index.html) emulate phenomena that are too small or too rapid to observe, such as chemical reactions or gas at the molecular level. These visual, interactive computational experiments for teaching and learning science can be customised and adapted by the teacher. Some science-learning environments have embedded formative assessments that teachers can access immediately in order to gauge the effectiveness of their instruction and modify their plans accordingly (Delgado & Krajcik, 2010).

Furthermore, a variety of recent educational games for science education could, in principle, integrate assessment and tutoring functionalities. ARIES (Acquiring Research Investigative and Evaluative Skills) is a computerised educational tool which incorporates multiple learning principles, such as testing effects, generation effects, and formative feedback (Wallace, et al., 2009). Another example is Quest Atlantis (http://atlantis.crlt.indiana.edu/), which promotes causal reasoning skills, subject knowledge in physics and chemistry, and an understanding of how systems work at both macro and micro level.

Some game environments include feedback, tutoring and monitoring of progress. In River City, for example, students use their knowledge of biology and the results of tests conducted online with equipment such as virtual microscopes to investigate the mechanisms through which a disease is spreading in a simulated 18th century city. Prompts gradually fade as students acquire inquiry skills. Data-mining allows teachers to document gains in students' engagement, learning and self-efficacy (Dede, 2010; Means & Rochelle, 2010).

Learner Response Systems

Another interesting example of how technology is being used for assessment purposes is the Learner Response Systems, often referred to as clickers. A recent study by the University of York (Sheard, Chambers, & Elliot, 2012) has looked into how clickers were being used and how they impacted on the teaching of grammar

in Year 5 classes across 42 schools in the North of England and Wales. It involved pupils using handsets to respond individually to question sets that were presented digitally. The formative assessment method used in this study is called Questions for Learning (QfL). The results indicate that, on grammar tests, students from QfL classes perform far better than those in the control classes. These effects did not, however, generalise to the writing task. An interesting finding is that middle- and low-performing students seem to benefit more particularly from formative assessment using handheld devices.

Authentic Tasks Using Digital Tools

Practical tasks using mobile devices or online resources are another promising avenue for developing ICT-enabled assessment formats. Several national pilots assess tasks that replicate real life contexts and are solved by using common technologies, such as the Internet, office and multimedia tools.

In Denmark, for example, students at commercial and technical upper secondary schools have, since 2001, been sitting Danish language, Maths and Business Economics exams based on CD-ROMs with access to multimedia resources. The aim is to evaluate their understanding of the subjects and ability to search, combine, analyse and synthesise information and work in an inter-disciplinary way (www.cisco.com/web/strategy/docs/education/DanishNationalAssessmentSystem.pdf.)

Similarly, in the eSCAPE project, a 6-hour collaborative design workshop replaced school examinations for 16-year-old students in Design and Technology in 11 schools across England. Students work individually, but within a group context, and record assessment evidence via a handheld device in a short multimedia portfolio. The reliability of the assessment method was reported as very high (Binkley *et al.*, 2012; Ripley, 2009b).

In the US, College Work and Readiness Assessment (CWRA) was introduced in St. Andrew's School in Delaware to test students' readiness for college and work, and it quickly spread to other schools across the US. It consists of a single 90-minute task that students must accomplish using a library of online documents. They must address real-world dilemmas (e.g. helping a town reduce pollution), making judgements that have economic, social, and environmental implications, and articulate a solution in writing.

The Key Stage 3 ICT tests (UK) require 14-year-old students to use multiple ICT tools in much the same way as in real work and academic environments (Bennett, 2010). Similarly, the iSkills (www.ets.org/iskills) assessment aims to measure students' critical thinking and problem-solving skills in a digital environment. In a one-hour exam real-time, scenario-based tasks are presented that measure the ability to navigate, critically evaluate and understand the wealth of information available through digital technology. The national ICT skills assessment programme in Australia (MCEECDYA, 2008) is designed to be an authentic performance assessment, mirroring students' typical 'real world' use of ICT. In 2005 and 2008, students completed tasks on computers using software that included a seamless combination of simulated and live applications.

These examples illustrate that technologies can be used to support authentic contexts and tasks that allow for a more comprehensive and valid assessment of 21st century skills, such as scientific inquiry, analysis, interpretation and reflection.

While this potential has not yet been exploited fully for assessment purposes in education and training, experimentation indicates that learning in real-life contexts can be promoted and more adequate and applied assessment strategies are supported.

The Conceptual Leap

As the examples outlined above illustrate, we are witnessing many innovative developments in the areas of Computer-Based Testing and embedded assessment and intelligent tutoring which offer promising avenues to capture complex key competences and 21st century skills. One could say that we are in a situation where the old 'Testing Paradigm' is reaching out to accommodate for the assessment needs of the 21st century, while we see a completely new way of learning and assessing emerging on the horizon, which is supported by technological advances that will need another few years to become mainstreamed.

In the past, research focused on the technological side, with more and more tools, functionalities and algorithms to increase measurement accuracy and create more complex and engaging learning environments with targeted feedback loops (Bennett, 2010; Bridgeman, 2009; Ridgway & McCusker, 2008). Given that learning analytics could, in the future, replace explicit testing, (e)Assessment will become far more closely interwoven with learning and teaching and will have to respond to and respect the pedagogical concepts on which the learning process is based. However, at the crossroads of the two learning and assessment paradigms, pedagogy is lagging behind in guiding technological innovation. With the 'old' testing paradigm, communication between pedagogy and technology was one-directional, with traditional assessment approaches and practical issues in their implementation guiding technological development. With the 'new' Embedded Assessment Paradigm, technological innovation is leading the way and pedagogy has not yet started to guide its course. To fully exploit the potential of new learning environments and embedded assessment, policy and pedagogy must reflect on concrete learning and assessment needs and enter into a dialogue with technology developers to ensure that newly emerging learning environments and tools adequately support 21st century learning. Research should therefore not only focus on increasing efficiency, validity and reliability of ICT enhanced assessment formats, but also consider how the pedagogical and conceptual foundations of different pedagogical approaches translate into different eAssessment strategies.

Furthermore, if data from the learning process itself can be used for assessment purposes in an objective, valid, reliable and comparable way, explicit testing will become obsolete. Thus, we need to re-consider the value of tests, examinations, and in particular of high stakes summative assessments that are based on the evaluation of students' performance displayed in just one instance, in dedicated tasks that are limited in scope. If these examinations do not change, there is a danger that educational practice and assessment will further diverge. If, on the other hand, formative and summative assessment become an integral part of the learning process, and digital learning environments become the main source for grading and certification, there is a need to better understand how information collected digitally should be used, evaluated and weighted to adequately reflect the performance of each individual learner. If genuinely pedagogical tasks, such as

assessing and tutoring, are increasingly delegated to digital environments, these must be designed in such a way that they become a tool for teachers and learners to communicate effectively with one another.

Hence, embedded assessment should be designed to respect and foster the primacy of pedagogy and the role of the teacher. Judgements on the achievement and performance of students that are based on data collected in digital environments must be based on a transparent and fair process of interpreting and evaluating these data which are mediated by digital applications and tools, but ultimately lie in the hands of teachers and learners. Since teachers will be able to base their pedagogical decisions and judgements on a wider range of data than in the past, pedagogical principles for interpreting, evaluating, weighing and reflecting on these different kinds of data are needed. Hence, it is important to ensure that progress in the technological development of environments, applications and tools for learning and assessment is guided by pedagogical principles that reflect the competence requirements of the 21st century.

Policy Options

Policy plays an important role in mediating change that can enable a paradigm shift in and for eAssessment which is embedded in the complex field of ICT in education. Policy is important in creating coherence between policy elements, for sense-making and for maintaining a longitudinal perspective. At the crossroads of two different eAssessment paradigms, moving from enhancing assessment strategies to transforming them, policy action and guidance are of crucial importance. In particular, seeing this leap as a case of technology-based school innovation (Johannessen & Pedró, 2010) the following policy options should be considered:

○ *Policy coherence*. One of the most important challenges for innovation policies for technology in education is to ensure sufficient policy coherence. The various policy elements cannot be regarded in isolation; they are elements that are interrelated and are often necessary for other policy elements to be effective. Policy coherence lies at the heart of the systemic approach to innovation through its focus on policy elements and their internal relations.

○ *Encourage research, monitoring and evaluation*. More research, monitoring and evaluation is needed to improve the coherence and availability of the emerging knowledge base related to technologies for assessment of and for learning. Monitoring and evaluation should, ideally, be embedded in early stages of R&D on new technologies and their potential for assessment. Both public and private stakeholders, e.g. the ICT industry can be involved in such initiatives.

○ *Set incentives for the development of ICT environments and tools that allow teachers to quickly, easily and flexibly create customised electronic learning and assessment environments*. Open source tools that can be adapted by teachers to fit their teaching style and their learners' needs should be better promoted. Teachers should be involved in the development of these tools and encouraged to further develop, expand, modify and amend these themselves.

○ *Encourage teachers to network and exchange good practice*. Many of the ICT-enhanced assessment practices in schools are promoted by a small number of teachers who enthusiastically and critically engage with ICT for assessment. To upscale and mainstream as well as to establish good practice, it is necessary to

better support these teachers, encourage them to exchange their experiences and establish good practice (See also Holmes in this issue pp. 97–112).

○ *Encourage discussion and guidance on viable ICT-enhanced assessment strategies.* While deployment of ICT in schools is lagging behind, given the vast range and variety of ICT strategies supporting assessment, a critical discourse on their advantages and drawbacks should be launched among educators and policy makers which can lead to recommendations for the take-up of ICT to support the comprehensive assessment of 21st century skills.

Conclusion

This article has argued for a paradigm shift in the use and deployment of Information and Communication Technologies (ICT) in assessment. In the past, eAssessment focused on increasing efficiency and effectiveness of test administration; improving the validity and reliability of test scores; and making a greater range of test formats susceptible to automatic scoring, with a view to simultaneously improve efficiency and validity. Despite the variety of computer-enhanced test formats, eAssessment strategies have been grounded in the traditional assessment paradigm, which has for centuries dominated formal education and training and is based on the explicit testing of knowledge.

However, against the background of rapidly changing skill requirements in a knowledge-based society, education and training systems in Europe are becoming increasingly aware that curricula and with them assessment strategies need to refocus on fostering more holistic 'Key Competences' and transversal or general skills, such as '21st century skills'. ICT offer many opportunities for supporting assessment formats that can capture complex skills and competences that are otherwise difficult to assess. To seize these opportunities, research and development in eAssessment and assessment in general must transcend the Testing Paradigm and develop new concepts of embedded, authentic and holistic assessment. Thus, while there is still a need to advance in the development of emerging technological solutions to support embedded assessment, such as Learning Analytics, and integrated assessment formats, the more pressing task is to make the conceptual shift between traditional and 21st century testing and develop (e-)Assessment pedagogies, frameworks, formats and approaches that reflect the core competences needed for life in the 21st century, supported by coherent policies for embedding and implementing eAssessment in daily educational practice.

Christine Redecker, Institute for Prospective Technological Studies (IPTS), European Commission, Joint Research Centre, Edificio Expo, C/ Inca Garcilaso 3, 41092 Seville, Spain, Christine.Redecker@ec.europa.eu

Øystein Johannessen, IT2ET Edu Lead, Qin AS, Gneisveien 18, N-1555 SON, Norway, oysteinjohannessen@educationimpact.net

Disclaimer

The views expressed in this article are purely those of the authors and may not in any circumstances be regarded as stating an official position of the European Commission.

NOTES

1. Examples include: Intelligent Essay Assessor (Pearson Knowledge Technologies), Intellimetric (Vantage), Project Essay Grade (PEG) (Measurement, Inc.), and e-rater (ETS).
2. For example: the Signals system at Purdue University (www.itap.purdue.edu/tlt/signals/); the Academic Early Alert and Retention System at Northerm Arizona University (www4.nau.edu/ua/GPS/student/).

REFERENCES

AMELUNG, M., KRIEGER, K. & RÖSNER, D. (2011) E-assessment as a service, *IEEE Transactions on Learning Technologies*, 4, pp. 162–174.

BARAB, S. A., SCOTT, B., SIYAHHAN, S., GOLDSTONE, R., INGRAM-GOBLE, A., ZUIKER, S. J. *et al.* (2009) Transformational play as a curricular scaffold: using videogames to support science education, *Journal of Science Education and Technology*, 18, pp. 305–320.

BEN-SIMON, A. & BENNETT, R. E. (2007) Toward a more substantively meaningful automated essay scoring; *Journal of of Technology, Learning and Assessment, 6.*

BENNETT, R. E. (2010) Technology for large-scale assessment, in: P. PETERSON, E. BAKER & B. MCGAW (Eds) *International Encyclopedia of Education* (3rd ed., Vol. 8, pp. 48–55) (Oxford, Elsevier).

BINKLEY, M., ERSTAD, O., HERMAN, J., RAIZEN, S., RIPLEY, M. & RUMBLE, M. (2012) Defining 21st century skills, in: P. GRIFFIN, B. MCGAW & E. CARE (Eds) *Assessment and Teaching of 21st Century Skills* (pp. 17–66) (Dordrecht, Heidelberg, London, New York: Springer).

BRIDGEMAN, B. (2009) Experiences from large-scale computer-based testing in the USA, in: F. SCHEUERMANN & J. BJÖRNSSON (Eds) *The Transition to Computer-Based Assessment* (Luxembourg, Office for Official Publications of the European Communities).

BUNDERSON, V. C., INOUYE, D. K. & OLSEN, J. B. (1989) The four generations of computerized educational measurement, in: R. L. LINN (Ed) *Educational Measurement* (Third ed., pp. 367–407) (New York, Macmillan).

BUTCHER, P. G. & JORDAN, S. E. (2010) A comparison of human and computer marking of short free-text student responses, *Computers and Education*, 55, pp. 489–499.

CACHIA, R., FERRARI, A., ALA-MUTKA, K. & PUNIE, Y. (2010) *Creative Learning and Innovative Teaching: Final Report on the Study on Creativity and Innovation in Education in EU Member States* (Seville, JRC-IPTS).

COALITION, I. E. (2012) *Enabling Technologies for Europe 2020.*

COUNCIL OF THE EUROPEAN UNION (2006) *Recommendation of the European Parliament and the Council of 18 December 2006 on key competences for lifelong learning*. (2006/962/EC) (Official Journal of the European Union, L394/10).

CSAPÓ, B., AINLEY, J., BENNETT, R., LATOUR, T. & LAW, N. (2010) *Technological Issues for Computer-Based Assessment*. http://atc21s.org/wp-content/uploads/2011/11/3-Technological-Issues.pdf

DE JONG, T. (2010) Technology supports for acquiring inquiry skills, in: P. PETERSON, E. BAKER & B. MCGAW (Eds) *International Encyclopedia of Education* (3rd ed., Vol. 8, pp. 167–171) (Oxford, Elsevier).

DEDE, C. P. (2010) Technological support for acquiring twenty-first-century skills, in: P. PETERSON, E. BAKER & B. MCGAW (Eds) *International Encyclopedia of Education* (3rd ed., Vol. 8, pp. 158–166) (Oxford, Elsevier).

DELGADO, C. & KRAJCIK, J. (2010) Technology and learning — supports for subject matter learning, in: P. PETERSON, E. BAKER & B. MCGAW (Eds) *International Encyclopedia of Education* (3rd ed., Vol. 8, pp. 197–203) (Oxford, Elsevier).

EGGEN, T. J. H. M. & STRAETMANS, G. J. J. M. (2009) Computerized adaptive testing of arithmetic at the entrance of primary school teacher training college, in: F. SCHEUERMANN & J. BJÖRNSSON (Eds) *The Transition to Computer-Based Assessment* (Luxembourg, Office for Official Publications of the European Communities).

EL-ALFY, E. S. M. & ABDEL-AAL, R. E. (2008) Construction and analysis of educational tests using abductive machine learning, *Computers and Education*, 51, pp. 1–16.

ELLIS, S. & BARRS, M. (2008) The assessment of creative learning, in: J. SEFTON-GREEN (Ed) *Creative Learning* (pp. 73–89) (London, Creative Partnerships).

ETS (2012) *Sea Change in Assessment: How Technology is Transforming K-12 Testing*: www.k12center.org/rsc/pdf/a-sea-change-in-assessment-letter-size.pdf.

EURYDICE (2012) *Developing Key Competences at School in Europe: challenges and opportunities for policy* (Education, Audiovisual and Culture Executive Agency).

FERRARI, A., CACHIA, R. & PUNIE, Y. (2009) *Innovation and Creativity in Education and Training in the EU Member States: Fostering Creative Learning and Supporting Innovative Teaching* (Seville, JRC-IPTS).

FLORIÁN, B. E., BALDIRIS, S. M., FABREGAT, R. & DE LA HOZ MANOTAS, A. (2010) *A set of software tools to build an author assessment package on Moodle: implementing the AEEA proposal*. 10th IEEE International Conference on Advanced Learning Technologies, ICALT (2010 ed., pp. 67–69).

FREY, A. & SEITZ, N. N. (2009) Multidimensional adaptive testing in educational and psychological measurement: current state and future challenges, *Studies in Educational Evaluation*, 35, pp. 89–94.

GARRETT, N., THOMS, B., ALRUSHIEDAT, N. & RYAN, T. (2009) Social ePortfolios as the new course management system, *On the Horizon*, 17, pp. 197–207.

GARTNER RESEARCH (2012) *Hype Cycle for Emerging Technologies*.

GRAFF, M., MAYER, P. & LEBENS, M. (2008) Evaluating a web based intelligent tutoring system for mathematics at German lower secondary schools, *Education and Information Technologies*, 13, pp. 221–230.

HARDRÉ, P. L., CROWSON, H. M., XIE, K. & LY, C. (2007) Testing differential effects of computer-based, web-based and paper-based administration of questionnaire research instruments, *British Journal of Educational Technology*, 38, pp. 5–22.

HARTIG, J. & HÖHLER, J. (2009) Multidimensional IRT models for the assessment of competencies, *Studies in Educational Evaluation*, 35, pp. 57–63.

HE, Y., HUI, S. C. & QUAN, T. T. (2009) Automatic summary assessment for intelligent tutoring systems, *Computers and Education*, 53, pp. 890–899.

HERRENKOHL, L. R., TASKER, T. & WHITE, B. (2011) Pedagogical practices to support classroom cultures of scientific inquiry, *Cognition and Instruction*, 29, pp. 1–44.

HORKAY, N., BENNETT, R. E., ALLEN, N., KAPLAN, B. & YAN, F. (2006) Does it matter if I take my writing test on computer? An empirical study of mode effects in NAEP, *Journal of Technology, Learning, and Assessment*, 5.

HUANG, C. J., WANG, Y. W., HUANG, T. H., CHEN, Y. C., CHEN, H. M. & CHANG, S. C. (2011) Performance evaluation of an online argumentation learning assistance agent, *Computers and Education*, 57, pp. 1270–1280.

HUTCHISON, D. (2007) An evaluation of computerised essay marking for national curriculum assessment in the UK for 11-year-olds, *British Journal of Educational Technology*, 38, pp. 977–989.

JOHANNESSEN, O. & PEDRÓ, F. (2010) Conclusions: lessons learnt and policy implications, in: OECD (Ed) *Inspired by Technology, Driven by Pedagogy: A Systemic Approach to Technology-Based School Innovations* (Paris, Centre for Educational Research and Innovation, OECD).

JOHNSON, L., ADAMS, S. & CUMMINS, M. (2012a) *The NMC Horizon Report: 2012 Higher Education Edition* (Austin, Texas, The New Media Consortium).

JOHNSON, L., ADAMS, S. & CUMMINS, M. (2012b) *NMC Horizon Report: 2012 K-12 Edition* (Austin, Texas, The New Media Consortium).

JOHNSON, L., SMITH, R., WILLIS, H., LEVINE, A. & HAYWOOD, K. (2011) *The 2011 Horizon Report* (Austin, Texas, The New Media Consortium).

JORDAN, S. & MITCHELL, T. (2009) e-Assessment for learning? The potential of short-answer free-text questions with tailored feedback, *British Journal of Educational Technology*, 40, pp. 371–385.

LENT, G. v. (2008) Important considerations in e-Assessment: an educational measurement perspective on identifying items for a European Research Agenda, in: F. SCHEUERMANN & A. G. PEREIRA (Eds) *Towards a Research Agenda on Computer-Based Assessment. Challenges and needs for European Educational Measurement* (Luxembourg: Office for Official Publications of the European Communities).

LJUNGDAHL, L. & PRESCOTT, A. (2009) Teachers' use of diagnostic testing to enhance students' literacy and numeracy learning, *International Journal of Learning*, 16, pp. 461–476.

LOONEY, J. (2010) Making it Happen: formative assessment and educational technologies. *Promethean Thinking Deeper Research Papers*, 1(3).

MARTIN, R. (2008) New possibilities and challenges for assessment through the use of technology; in: F. SCHEUERMANN & A. G. PEREIRA (Eds) *Towards a Research Agenda on Computer-Based Assessment* (Luxembourg: Office for Official Publications of the European Communities).

MCEECDYA (2008) *National Assessment Program. ICT Literacy Years 6 and 10 Report 2008*. Australia: Ministerial Council for Education, Early Childhood Development and Youth Affairs (MCEECDYA).

MEANS, B. & ROCHELLE, J. (2010) An overview of technology and learning, in P. PETERSON, E. BAKER & B. MCGAW (Eds) *International Encyclopedia of Education* (3rd ed., Vol. 8, pp. 1–10) (Oxford, Elsevier).

MOE, E. (2009) Introducing large-scale computerised assessment lessons learned and future challenges, in: F. SCHEUERMANN & J. BJÖRNSSON (Eds) *The Transition to Computer-Based Assessment* (Luxembourg, Office for Official Publications of the European Communities).

NACCCE (1999) *All Our Futures: creativity, culture and education.*

NICOL, D. & MACFARLANE-DICK, D. (2006) Formative assessment and selfregulated learning: A model and seven principles of good feedback practice, *Studies in Higher Education*, 31, pp. 199–218.

NOORBEHBAHANI, F. & KARDAN, A. A. (2011) The automatic assessment of free text answers using a modified BLEU algorithm; *Computers and Education*, 56, pp. 337–345.

NUNAN, D. (2010) Technology Supports for Second Language Learning, *International Encyclopedia of Education* (3rd ed.,Vol. 8, pp. 204–209) (Oxford, Elsevier).

OECD (2010) *The Nature of Learning: using research to inspire practice* (Paris, OECD).

PARK, J. (2010) Constructive multiple-choice testing system, *British Journal of Educational Technology*, 41, pp. 1054–1064.

PELLEGRINO, J. W. (2010) Technology and learning — assessment, in P. PETERSON, E. BAKER & B. MCGAW (Eds) *International Encyclopedia of Education* (3rd ed.,Vol. 8, pp. 42–47) (Oxford, Elsevier).

PUENTEDURA, R. (2012) *Building Upon SAMR*: www.hippasus.com/rrpweblog/archives/2012/09/03/BuildingUponSAMR.pdf.

REDECKER, C., LEIS, M., LEENDERTSE, M., PUNIE, Y., GIJSBERS, G., KIRSCHNER, P. *et al.* (2010) *The Future of Learning: preparing for change*. (Seville, JRC-IPTS).

RIDGWAY, J. & MCCUSKER, S. (2008) Challenges for research in e-Assessment, in: F. SCHEUERMANN & A. G. PEREIRA (Eds) *Towards a Research Agenda on Computer-Based Assessment* (Luxembourg: Office for Official Publications of the European Communities).

RIPLEY, M. (2009a) Transformational computer-based testing, in: F. SCHEUERMANN & J. BJÖRNSSON (Eds) *The Transition to Computer-Based Assessment* (Luxembourg: Office for Official Publications of the European Communities).

RIPLEY, M. (2009b) Transformational computer-based testing, in: F. SCHEUERMANN & J. BJÖRNSSON (Eds) *The Transition to Computer-Based Assessment. New Approaches to Skills Assessment and Implications for Large-scale Testing* (Luxembourg, Office for Official Publications of the European Communities).

RITTER, S., ANDERSON, J. R., KOEDINGER, K. R. & CORBETT, A. (2007) Cognitive tutor: applied research in mathematics education, *Psychonomic Bulletin and Review*, 14, pp. 249–255.

RITTER, S., TOWLE, B., MURRAY, R. C., HAUSMANN, R. G. M. & CONNELLY, J. (2010) *10th International Conference on Intelligent Tutoring Systems, ITS 2010: Vol. 6095 LNCS* (pp. 452) Pittsburgh, PA.

SADLER, D. R. (1989) Formative assessment and the design of instructional systems. *Instructional Science*, 18, pp. 119–144.

SHEARD, M., CHAMBERS, B. & ELLIOT, L. (2012) *Effects of Technology-enhanced Formative Assessment on Achievement in Primary Grammar* (York, UK, The University of York).

THOMPSON, N. A. & WEISS, D. J. (2009) Computerized and adaptive testing in educational assessment, in: F. SCHEUERMANN & J. BJÖRNSSON (Eds) *The Transition to Computer-Based Assessment* (Luxembourg: Office for Official Publications of the European Communities).

WALLACE, P., GRAESSER, A., MILLIS, K., HALPERN, D., CAI, Z., BRITT, M. A. *et al.* (2009) Operation ARIES!: A computerized game for teaching scientific inquiry, *Frontiers in Artificial Intelligence and Applications*, 200, pp. 602–604.

WANDALL, J. (2009) National tests in Denmark — CAT as a pedagogic tool, in: F. SCHEUERMANN & J. BJÖRNSSON (Eds) *The Transition to Computer-Based Assessment* (Luxembourg, Office for Official Publications of the European Communities).

WEIGLE, S. C. (2010) Validation of automated scores of TOEFL iBT tasks against non-test indicators of writing ability, *Language Testing*, 27, pp. 335–353.