COMPUTER-BASED EXAMS IN SCHOOLS: FREEDOM FROM THE LIMITATIONS OF PAPER?

C. PAUL NEWHOUSE

Centre for Schooling and Learning Technologies, Edith Cowan University, 2 Bradford St, Mount Lawley, Western Australia 6050
p.newhouse@ecu.edu.au

There is little doubt that the curriculum content and pedagogy in schools is driven by the structure and forms of assessment employed, particularly for summative purposes. When most such assessment was limited to what a student could do with a pen and paper in short 'exams' this pushed the content of the curriculum towards small descriptive chunks, and the pedagogy towards memorisation and replication techniques. Over the past two decades alternative forms of assessment supported by the power of computer systems have been conceived and tried. This paper discusses progress towards various forms of computer-based exams and how these may encourage curriculum and pedagogy suited to 21st Century learning. In particular it draws attention to research I have led to investigate the feasibility of a number of forms of computer-based exams for high-stakes summative assessment in secondary education. The result was the successful development and implementation of audiovisual stimuli and response computer-based exams, and digital production exams resulting in portfolios of evidence. Each implemented form of exam was tested for feasibility to ensure a defensible balance of manageability, reliability and validity. I believe that for most high-stakes summative assessment in Australian schools it is time to replace the ancient paper-based technology with computer-based technologies.

Keywords: Computer-based exams; summative assessment.

1. Introduction

The history of examinations begins over 1000 years ago with origins believed to be in China where written examinations were developed as part of a promotions system for the Imperial Civil Service (Stobart, 2008, p. 102). It is therefore perhaps surprising that most secondary school children still endure this type of assessment regime in 21st Century Australia. The image of children sitting in rows writing on paper for two or three hours still seems to be the epitome of the outcomes of more than 12 years of schooling. For many students these exam events will be the rare times in their lives they will use this ancient technology in such a prolonged manner (Way, Davis, & Straine-Seymour, 2008). The constraints of this form of assessment leads to the measurement of very limited outcomes often described as memorisation, repetition, or recall (Stobart, 2008). These contrast with the often stated required outcomes of schooling that include higher-order thinking skills, problem-solving, communication and collaboration (Clarke-Midura & Dede, 2010). There are some that still see over-riding benefits in paper-based assessment (e.g. Llamas-Nistal, Fernández-Iglesias, González-Tato, & Mikic-Fonte, 2013) but many

(e.g. Clarke-Midura & Dede, 2010; Dochy, 2009) call for new assessment modes that are characterised by students constructing knowledge, the application of this knowledge to real life problems, the use of multiple perspectives and context sensitivity, active participation, and the integration of assessment and the learning environment. Why isn't this more prevalent after 1000 years?

Stobart (2008) claims that every assessment is a "trade-off of construct validity (the domain or skill being tested), reliability and manageability" (p. 4) and the higher the stakes the greater the emphasis on the latter two at the expense of construct validity. He illustrates how in traditional paper-based exams particular constructs in a domain of learning may be omitted because they are too difficult to assess. Further, students may be taught how to pass an exam despite never having "learned the concepts on which they are being tested" (p. 6). Either way the assessment lacks authenticity, or what he calls dependability, and indicates the need to look for alternative forms of assessment. Taylor (2005) raises a concern about fairness of assessment based on hand-writing because, "… as increasing numbers of students are taught the writing process on computers, to require them to demonstrate writing skills on a paper and pencil test is a questionable endeavour. As a result, measures of student writing abilities in these cases may be substantially underestimated" (p. 95). These arguments concerning the inadequate validity of paper-based assessments in terms of what is intended to be measured, what needs to be measured for future use, and the technology of assessment (e.g. hand-writing) are also presented by many others (e.g. Bennett, 2002; Clarke-Midura & Dede, 2010; Pellegrino & Quellmalz, 2011; Way et al., 2008).

The Joint Research Centre for the European Commission (Scheuermann & Bojornsson, 2009) commissioned a major report titled: *The Transition to Computer-Based Assessment*. Within this report Kozma (2009) lays out a rationale for computer-based assessment in terms of a mismatch between what is needed in modern society and what is addressed, and thus assessed at school. In particular he draws attention to the differences between standardized pen-and-paper assessment and "tasks in the outside world" that are based on "complex ill-structured problems" that are solved "collaboratively" using "technological tools". While he does not see assessment reform only requiring the use of digital technologies he outlines a number of significant advantages including: reduced costs; increased adaptability to individuals; opportunity to collect process data on student performance; the provision to tools integral to modern practice; and better feedback data. Kozma does introduce a number of challenges to using digital technologies including: start-up costs for systems; the need to choose between standardized and 'native' applications; the need to integrate applications and systems; the need to choose between 'stand-alone' and online implementation; the need for security of data; and limited examples of high-quality assessments supported by digital technologies. He also highlights methodological challenges including: the extent of equivalence with pen-and-paper; the design of appropriate complex tasks; making efficient and reliable high-level professional judgements; scoring students' processes and strategies; and distinguishing individual contributions to collaborative work.

Clarke-Midura and Dede (2010) support the arguments of Kozma and claim that, "Paper-and-pencil tests are barely adequate to measure the minimum competencies required for low-level roles in industrial settings and fall woefully short of providing measures of the sophisticated knowledge and skills students need for 21st-century work and citizenship" (p. 312). But they do see, "Current technological advances offer exciting opportunities to design assessments that are active and situative, and that measure complex student knowledge and provide rich observations for student learning" (p. 311) and that were "… previously impossible to test in paper-based or hands-on formats" (p. 315). These opportunities are being explored by an ongoing research initiative of Cisco, Intel, and Microsoft (2009) referred to as the *Assessment and Teaching of 21st Century Skills* project. Their initial paper clearly argues that the focus of outcomes in schools needs changing to meet the needs of 21st Century but this change will not occur without changes to high-stakes assessment.

> *Reform is particularly needed in education assessment-how it is that education and society more generally measure the competencies and skills that are needed for productive, creative workers and citizens. … accountability efforts have measured what is easiest to measure, rather than what is most important. …. New assessments are required that measure these skills … and support systemic education reform. … assessments should engage students in the use of technological tools and digital resources and the application of a deep understanding of subject knowledge to solve complex, real world tasks and create new ideas, content, and knowledge. (Cisco et al., 2009, p. 1)*

Lesgold (2009) agrees with the rationale for computer-based assessment but identifies the initial step as developing a public "shared understanding" about what should be the outcomes from schooling and thus linking to resultant changes to assessment. He is confident 21st century skills cannot be assessed by the traditional standard approach to testing based on responses to small items that minimises the need for human judgement in marking. Instead students will need to tackle complex tasks using technological tools with experts judging the performances. Pellegrino and Quellmalz (2011) describe a large range of examples where the "computer's ability to capture student inputs permits collecting evidence" (p. 120) of many types upon which to assess "levels of expertise" using modern psychometric techniques. This will not be easy to accomplish and will need research, development and training.

At this stage it is likely that few teachers in Australian schools provide students with experiences in digital forms of assessment. For example, in a survey reported by Becta (2010) it was found that at best 4 out of 10 British teachers reported using the technologies to just 'create or administer assessment'. There is no reason to believe the situation in Australia would be any different. This lack of experience for students and teachers is likely to be a constraint in using computers to support summative assessment,

particularly where the stakes are high. However, at the same time non-school organisations are increasingly embracing computer-based testing. Even back in 2005 Taylor (2005) reported that, "Nearly a million computer-based examinations are delivered monthly throughout the world in a secure test centre environment" (p. 93). He points out that for organisations looking to convert from paper-based assessment the costs had reduced particularly because much of the infrastructure is already in place (e.g. hardware, software and networking). If all this was the case in 2005 how much more compelling are the arguments in 2013.

2. What are Computer-Based Exams?

From the early years of educational computing in the 1960s it was envisaged that exams or tests could be done on computers, but typically this only considered doing on the computer what was already done on paper. That is, the computer would present questions to which students would type in a response either by selecting from a list of options or typing extended text (Bull & Sharp, 2000). Over the years other ways in which assessments could be facilitated by using computers have been envisaged and in many cases tried (Clarke-Midura & Dede, 2010; Pellegrino & Quellmalz, 2011). Pilots and surgeons often do computer-based exams using simulators, some oral language exams have included 'talking' into the computer (Pearson Education Australia, 2012), drawing and audiovisual digital tools have been used to present and record performances, and the e-scape project in the UK had students using portable computers to record their progress working on practical work (Kimbell, Wheeler, Miller, & Pollitt, 2007). In short, an exam should measure some sort of performance and there is a multitude of ways in which this could be done on a computer system using the range of hardware, software and networking options now available. For example, a computer-based exam could be delivered on a stand-alone personal computer, or within an isolated Local Area Network (LAN) or use online technologies such as web-pages over the Internet (Taylor, 2005). Technical and administrative methods can be used to assure security and fairness (McNulty et al., 2011). The increasingly sophisticated range of peripheral devices can provide audiovisual and kinaesthetic support to collect a large range of types of student responses and evidence of performance.

For more than a decade researchers at the Centre for Schooling and Learning Technologies (CSaLT) at Edith Cowan University (ECU) have grappled with the multitude of ways in which digital technologies can be used to improve assessment, including with computer-based exams (Newhouse & Njiru, 2009). In so doing we have identified a number of categories of exams or exam items. Traditional exams are based on response items but an exam could include performance tasks or be based on the production of an artefact(s). Computer-based response items can provide stimuli using a full range of sensory forms and can allow responses in text, graphic, audio and kinaesthetic forms. Performance task items may involve recording a representation of a performance (e.g. making an audio recording of a talk), simulating a performance on the computer (e.g. demonstrating capability in using virtual equipment), or in some cases the

performance is in using software on the computer (e.g. an ICT literacy test). Similarly production items may involve recording the process of production and/or representing the final artefact(s) (e.g. e-scape project (Kimbell et al., 2007)). Naturally in practice the distinction between these categories of items is not always clear and there may be some blending.

What is the difference between a computer-based exam and any learning activity done on computer? In reality very little, except that the outputs from the exam are to be judged to measure knowledge and/or skill and therefore a set of constraints or conditions are typically imposed such as a length of time, access to resources (including persons), and form of outputs. These are imposed to assist in the process of measurement to enhance reliability, validity, and fairness (Bennett, 2002). However, when compared with paper and pen technology, digital technologies afford both a greater range, and greater control over the setting of conditions for exams (McNulty et al., 2011). For example, it is easier to set variations in time restrictions both in when and how long. Similarly variations in access to resources and forms of outputs and other such conditions are easier to allocate and monitor (e.g. access to the Internet, certain sites or particular documents). For example, Fluck, Pullen, and Harper (2009) conducted research in which computer-based exams were 'booted' off optical discs or USB flash drives to ensure students could not access the Internet or other resources not permitted for the exam.

There have been trials of various forms of computer-based exams in a number of countries including the UK, Norway, Denmark, USA and Australia. In Australia a computer-based test was used to assess the computer literacy of 12 and 16 year-old students (Ministerial Council on Education, Employment, Training and Youth Affairs, 2007). The test was developed around a simulated computer environment and implemented using sets of networked laptop computers. While the test was successfully implemented with over 7000 students this was over a long period of time and would not be readily scalable. Also the use of a simulated environment would be expensive and difficult to provide a great enough variety of activities each year. The trial in the UK also involved a multi-million pound simulated system but was accessed by students through their school computers (Boyle, 2006). In the Norwegian example students used their own government-provided notebook computers (British Broadcasting Corporation, 2009). In the USA increasingly national testing is computer-based and computer literacy will be included within national testing as is already the case in a number of states (Clarke-Midura & Dede, 2010; Harris, 2008).

Computer-based exams are a part of e-assessment that Ripley (2009) defines as "the use of technology to digitise, make more efficient, redesign or transform assessments and tests" that includes "professional examinations, qualifications, certifications and school tests, classroom assessment and assessment for learning" (p. 93). He proposes that there are two "drivers" of e-assessment being "business efficiency" and "educational transformation". E-assessment driven by business efficiency will tend to use digital technologies to support the same forms of assessment traditionally done on paper such as multiple-choice and short answer items. Where the driver is educational transformation

then it is likely that the e-assessment will take on different forms and designs to suit the outcomes or the performances to be assessed. He raises obstacles to be addressed similar to Lesgold (2009) but includes finding a method of scoring that is publicly acceptable and addressing the perceptions and conceptions of stakeholders about the technology and alternative forms of assessment. These obstacles provided a backdrop for our research into using digital technologies to support high-stakes summative assessment, within which a number of forms of computer-based exams were investigated.

3. Our Research with Computer-Based Exams

My first use of computer-based exams was as a secondary teacher in the early 1980s when teaching 'computing' and assessing performance tasks on the computer by collecting files created by students or observing them using the software. This included assessing the competence of some students in the use of one of the early flight simulators. In the early 1990s I worked with colleagues to develop interactive multimedia computer-based response exams allowing tertiary students to drag images to respond to mathematics questions and simulate the operation of safety devices on drilling platforms. In the early 2000s we used portable devices to record proficiency of tertiary students completing tasks on computers and used online technologies to collect digital artefacts created by students (Newhouse, 2006). Then in the mid-2000s we decided to tackle the high-stakes assessment at the end of schooling that had stubbornly retained traditional paper-based exams. Some of the detailed results from this research have been published (Newhouse, 2011; Penney, Jones, Newhouse, & Campbell, 2012). This paper provides a synthesis of those results in order to discuss their implications.

As humans we develop technologies primarily to address problems and increase productivity; the same should be true for using technologies to support assessment. It could be argued that digital technologies are used to support assessment either to save money (i.e. increase productivity) or address validity or reliability problems (i.e. improve the authenticity and measurement accuracy) (Ripley, 2009). If only simple multiple-choice and short answer exam items are considered then it is difficult to get anything cheaper than paper, pencil and 'granny' invigilators. Therefore our research focussed on areas where practical performances should be assessed and could not be validly or reliably assessed using simple multiple-choice and short answer exam items. The four courses we investigated each had a different form of practical performance to assess that gave rise to a rationale for considering the use of digital technologies. A summary of these for the four courses is provided in Table 1.

Table 1: The four courses investigated for digital forms of assessment.

Course	Nature of performance	Existing final assessment	Rationale for digital forms of assessment
Applied Information Technology (AIT)	Use digital technologies to create solutions to problems or challenges.	Written exam with multiple-choice, short answer and extended response items.	Increase validity by assessing a performance in which students use digital technologies.
Engineering Studies (ES)	Use technology and science knowledge to design and build prototypes of solutions to engineering challenges.	Written exam with multiple-choice, short answer and extended response items.	Increase validity and reliability by assessing an audiovisual digital recording of a performance and evidence.
Italian Studies (IS)	Oral and aural language proficiency.	Face-to-face interview with examiners.	Reduce costs and increase reliability by assessing an audiovisual digital recording of a performance or simulated performance.
Physical Education Studies (PES)	Apply sporting skills/ techniques and knowledge to respond to tactical challenges.	Face-to-face live skill drills and game scenarios in front of examiners.	Reduce costs and increase validity and reliability by assessing an audiovisual digital recording of a performance.

For both AIT and ES at the time of the research the final assessment was a written exam and yet for both courses the syllabus outlined an intent that students learn through practical application of knowledge and skills. For AIT the use of digital technology was central to the content, that is, it was intended that students use the technologies to learn about the technologies. For ES the content included practical design and prototyping using materials that would be expensive to assess through face-to-face observation. For IS and PES there were face-to-face final assessments but these were expensive, required students to travel to a central location, and because judgements were made on-the-spot were prone to unreliability as measures of performance, a problem recognised elsewhere in similar situations (Weir & Connor, 2009). Therefore for these latter two courses there was potential to use digital technologies to allow recordings of performances to be assessed at a later time with a view to decreasing costs, increasing reliability, and/or increasing validity.

In our study a variety of forms of computer-based exams were tried in all four courses as an assessment, while some other forms of assessment, such as portfolios, were also tried. For AIT a combination of production and performance tasks computer-based exam was used. For the other three courses audiovisual recording of production and/or performance activity combined with computer-based response items were used. The development and evaluation of these forms of assessment occurred in three annual cycles with different samples of students and teachers (some teachers were in more than one cycle). The three cycles were referred to as the 'proof of concept' phase, the 'prototype' phase and the 'scale-up' phase. Therefore in general for each course a similar form of

assessment was used in each phase with small improvements made from the evaluation in the previous phase except for in IS where there were some major changes in each phase.

The AIT exam involved students responding to a given challenge in the form of a design brief, by following a guide to develop a prototype digital solution over a two-hour period. For example, in the third year students designed and created a help-system for children to learn how to use a smartcard with the local public transport system. They were guided to use a standard personal computer to present a design, edit provided graphics and audiovisuals, create a prototype and evaluate the process and product. All work was submitted in digital files on USB flash drives, that also contained the digital materials to be used, and through an online system to be marked externally.

The ES exam was based on the model provided by the e-scape project in the UK (Kimbell et al., 2007). Students were given a challenge in the form of a design brief and were then guided through on-screen instructions to present a design and create a physical prototype using provided materials. At various times throughout the exam they were required to take photographs, audio and video recordings, and type and/or draw responses to questions to represent the development of their production. The system automatically collated these into a portfolio of evidence to be marked externally.

The PES exam involved students responding to a sporting tactical challenge, related to their chosen sport, and being guided to respond to on-screen questions and audiovisual recordings of their involvement with skill drills and game scenarios. They used text and drawing software through a database to give theoretical responses to questions about the challenge, they were then video-recorded completing skills drills and game scenarios related to the challenge, and finally viewed their videos and responded to review and evaluation questions using text and drawing software. Video recording was accomplished using a remotely controlled combined-feed multi-camera system that outputs a single digital file and game scenarios ensured each student was presented with the tactical challenge. These files along with database files containing text and drawing responses were marked externally.

The IS exam included students listening and/or watching a conversation and audio-recording a response on a personal computer system. In the final year a video was created with actors speaking in Italian that included an invitation to 'join the conversation'. This was designed to assess listening and speaking skills. In addition, in the final year a set of computer-based multiple-choice and short answer response items were added. All student audio recordings were captured in digital files to be marked externally.

For all four courses the resulting digital files were accessed externally by assessors using software systems operated through standard Internet browsers. Two or three assessors judged the performance of each student against analytical criteria represented in a rubric. As a result standards-referenced scores were generated using a Rasch polytomous model (Bennett, Tognolini, & Pickering, 2012). In addition for each course the set of performances was assessed using a comparative pairs method employing around 10 assessors for each course. This method required assessors to select 'winners' between pairs of performances with all of the resulting data being used to generate scores

Table 2: Summary of samples and cases for the three years of the study.

Course	Year of study	Cases	School type		# Teachers	# Students
			Public	Private		
AIT	1	7	2	5	7	115
AIT	2	6	4	2	6	85
AIT	3	16	9	7	16	163
ES	1	5	4	1	6	66
ES	2	5	3	2	6	78
ES	3	8	3	5	8	94
IS	1	4	2	2	4	35
IS	2	4	0	4	4	64
IS	3	7	0	7	7	97
PES	1	4	2	1	3	39
PES	2	4	2	0	3	27
PES	3	11	4	2	12	152
TOTALS		81	35	38	82	1015

using a Rasch dichotomous model. Research by Heldsinger and Humphry (2010) has found this method to yield scores with "high reliability and concurrent validity" (p. 17). The two methods of marking were used to investigate the relative reliabilities of the forms of assessment. Table 2 provides a summary of the samples representing the scope of the study.

4. Findings from Our Research

Some of the findings of our research are now summarised in terms of three forms of exam: production; performance tasks; and digital response.

4.1. *Production exams*

Students in AIT were assessed in a computer-based exam that involved a limited production. For ES a form of computer-supported production exam was used to guide the construction of an 'evidence' portfolio. These exams were generally well accepted by students and teachers as being well aligned with their pedagogical expectations or intentions. Production exams were more easily managed and facilitated with digital technologies where local storage (e.g. USB flash drives) was used, as was always the case for AIT. For ES, where firewalls could be negotiated, online exams posed few problems and were easy to manage. For both these forms of exam it was demonstrated that they could be implemented relatively inexpensively and could be reliably marked externally online.

4.2. *Performance tasks exams*

Students in AIT were assessed in a computer-based exam that included some performance task items. For IS forms of computer-based performance exam items were

used to assess speaking, listening and viewing and in PES performance exam items were video-recorded. The performance aspect of the AIT exam provided some difficulties in students knowing what software to access and with some software operating poorly on school hardware. However, they were more likely to enjoy, and engage positively with, a performance-based digital assessment task as opposed to a written examination.

For PES the video recording of sporting activities performance was relatively easily accomplished using the camera system that could be operated by anyone with less than an hour of training. Although the camera system would cost nearly $10K, less sophisticated cheaper systems would probably be adequate. Students and teachers accepted that judging a recording of the performance was more reliable than on-the-spot judging. They also perceived that being able to review and evaluate performance by viewing the videos increased the validity of the assessment.

For IS computer-based exams were more easily managed and facilitated where local storage (e.g. USB flash drives) was used. Students engaged positively but generally they and their teachers preferred a face-to-face interview to an audio-recorded performance. This preference appeared to stem from a perception that the former was more reliable as a measure and the increased flexibility increased its validity. However, the study found that scores from face-to-face interviews were not more reliable than those from audiovisual-recordings of the performances and that scores were very similar from both.

4.3. *Digital response exams*

Students in PES were assessed in two components of the task using typed and digitally drawn responses to questions and in ES various items of the exam using typed and digitally drawn responses. For IS both multiple-choice and typed response items were used in the third year. For AIT in the first year students responded to questions using typed responses as a component of the exam. Digital response exams provided few difficulties and were easy to manage, particularly where local storage was used, such the USB flash drives used in PES and AIT. Online digital response exams involving multiple-choice and typed responses posed very few difficulties other than ensuring access to the system through the school firewall. In a number of schools this was not possible using the e-Scape system used in ES but was less problematic in IS using the Willock online exam system (A. Willock Information Systems, 2013). However, the latter system proved unreliable at some sites in not uploading all audio recordings. The inclusion of drawing tools in PES and ES were successful, added critical functionality, and generally were handled well by students.

5. Discussion

In 2005 Taylor (2005) envisaged that technology would "… bring with it a host of benefits that go far beyond the capability of a paper-and-pencil testing environment" (p. 102). He listed potential advantages that I have compared with the findings from our research in Table 3. Apart from the use of item banks that we did not investigate, all the potential benefits envisaged by Taylor were realised to some extent in our study. In

Table 3: Potential advantages of digital forms of assessment over paper-based forms.

Potential advantage envisaged by Taylor (2005, p. 102)	Findings from our research
(1) A closer match between curriculum and instruction by enhancement of item types – simulations, models, sound, etc.	Students and teachers recognised in all courses that digital technologies allowed assessment tasks to provide a closer alignment with preferred pedagogies and curriculum intentions (Newhouse, 2012).
(2) More extensive use of existing banks of items.	Not investigated.
(3) Greater precision of measures through capacity to adapt to individual student competency levels.	Reliable measurement could occur in all digital forms of assessment that were tried (Newhouse, 2011, 2012). Rasch analysis of both analytical and comparative pairs marking yielded reliability coefficients in excess of 0.88 for each course (Newhouse, 2012).
(4) Ability to measure learning outcomes not possible through paper and pencil.	Most students and teachers recognised that digital technologies were used to support assessment tasks that measured practical performance that was not possible using paper-based assessments (Newhouse, 2012).
(5) Cost savings and increased reliability in marking.	Marking processes were efficient, cost-neutral and generated reliable sets of scores. The comparative pairs method of marking appeared to be superior where there was a main outcome from a fairly open-ended task (Newhouse, 2011).
(6) Greater access for students through the potential for examination on demand.	Online versions of exams tried in ES and IS were on demand but varied in success depending on school network behaviour (Newhouse & Cooper, 2013; Pagram & Williams, 2012).
(7) Vastly improved turnaround time to provide students with instant personalized feedback and teachers with information for remediation and instruction.	In the second year students in IS were provided with feedback through the online system used to deliver the exam (Newhouse & Cooper, 2013).
(8) Savings in shipping, handling, and printing costs.	These savings were particularly evident in AIT where cheap USB flash drives were used to deliver and collect the exam (Newhouse, 2013).
(9) Increased instructional time by reducing labour-intensive marking activities.	Marking using online digital tools was more efficient than paper-based systems (Newhouse, 2011).
(10) Increased student ownership of learning through tools to increase their involvement, control and motivation.	This was somewhat evident in AIT and ES where students had choice of prototype, and in IS where multiple attempts at oral recordings were permitted (Newhouse, 2012, 2013; Newhouse & Cooper, 2013).

particular the focus of our research was on more closely matching assessment and the curriculum, measuring learning outcomes not possible using paper-based methods, increasing the precision of measurement and ensuring cost-effectiveness in typical school

environments. Each of these was demonstrated in the form of assessment used in at least one of the four courses as described in Table 3.

Since we began our main study others in Australia, and beyond, have advanced our knowledge of the use of computer-based exams (Clarke-Midura & Dede, 2010). Some have focussed on a particular technology for assessment while others have focussed on either cost-effectiveness or improving the assessment of particular learning outcomes (Pellegrino & Quellmalz, 2011). For example, in Australia Fluck et al. (2009) developed an eExamination system with very strong controls including the option of running a completely locked up local environment. Their research was with tertiary students and generally just provided simple response items with students typing in a word processed document. Panizzon, Elliott, and Westwell (2010) successfully used a similar system run off USB flash drives to deliver multimedia rich science response exams. In NSW a project was also implemented in science assessment to not only use multimedia animations to assess science processes but also reduce the costs by using a large online system (NSW Department of Education and Training, 2013). Overseas we have seen a similar focus in the American assessment of writing for national testing with 52,200 students across the nation (National Center for Education Statistics, 2012).

In line with our research findings most researchers have found that schools in countries such as Australia typically have adequate technical infrastructure to implement such computer-based exams. However, as Panizzon et al. (2010) found there can still be some difficulties with access to workstations and navigating school constraints such as not being able to open an .exe file, or save to a USB. Despite this the factors inhibiting the use of computer-based exams in schools now are not so much technical rather concern the perceptions of students, teachers, school administrators and the public. While Fluck et al. (2009) were able to readily implement computer-based exams using their eExamination system they found that only about half the students preferred this to written exams. They recognised that this was at odds with the findings of other researchers. However, in agreement with our findings they did find that "students who had previously taken a computer based exam preferred this medium (63% of respondents) compared with 37% of first-timers" (p. 516). In our research we found particular groups of teachers wanted to 'hang on to tradition' and perceived unsubstantiated limitations with digital systems (e.g. Italian Studies teachers). In general Stobart (2010) argues that teachers tend to want "predictable" (p. 9) tests which mitigates against change in assessment policies and practices.

Change in the structure of schooling tends to occur incrementally (Stobart, 2010) and it would appear that this is particularly the case for assessment policies and practices. This is probably because quantum change requires support from teachers, school leaders, parents and the wider community. Typically such change requires political decisions by government and school system leaders concerning policies (validity and reliability concerns) and resource allocations (manageability concerns) and in democratic communities that means support from the wider community (Stobart, 2010). Dede (2008) sees this political factor as a major barrier to the use of digital technologies in schools,

not only for assessment purposes; the perceptions of the wider community of the nature of assessment and learning itself need changing.

Unfortunately the wider community tends to have entrenched views about schooling based on previous experience of schooling that everyone has had. This is particularly so for perceptions of the nature of assessment with Stobart (2010) claiming that testing has shaped our identities as learners, perhaps represented in the attitude of, 'it was good enough for me'. There tends to be a community perception that the only way to assess whether a student knows something is if they can answer questions on paper, in a short period of time, by themselves, and that if it is marked anonymously the aggregated scores represent capability. It has amused me when people tell me that they have never used anything they learned at school or University and yet in the next breath say that they want children to go through the same system. Then there is the advertisement that was on radio in our state for a maths tuition system that reminded parents that they feel that they can't help their children because they have forgotten it all from their school days. So clearly changing community perceptions is going to be difficult, but is necessary if schooling is to be both engaging and relevant.

6. Conclusion

It is generally recommended that computer-based exams will be successful where the tasks are clearly defined and limited, the work environment is prescribed (e.g. access to prescribed information or tools), and the required work output is well defined. The legitimate areas of concern are ensuring fairness for all students (e.g. access to information and technologies), the validity of the assessment, and reliably marking the typically varied types of output. Our research set out to achieve these outcomes for a diverse range of forms of computer-based exams to assess a variety of types of learning outcomes associated with different subject disciplines. We have demonstrated the feasibility of using digital technologies to support this endeavour for high-stakes summative assessment. This adds to the plethora of trials and research recently conducted around the world that provides a compelling argument for the replacement of paper-based exams with computer-based exams and other digital forms of assessment.

A computer-exam may be an enhanced version of a paper-based exam with multiple-choice, short answer and extended writing in the way that has been demonstrated by science exams in NSW system (NSW Department of Education and Training, 2013), the writing test for NAEP in the USA (National Center for Education Statistics, 2012) and Fluck's eExaminations (Fluck et al., 2009). Multimedia stimulus material can be added, interactive objects can be manipulated, and adaptivity can be added as illustrated in our PES exam. We have demonstrated that a computer-based exam can also be devised to manage a practical production and collect evidence of performance, as for our ES course. Students can create and submit digital artefacts as part of a computer-based exam, as for our AIT exam, or can record audiovisual responses as was done in our IS exam. These can all be readily implemented in our schools and increasingly can be delivered and monitored online. There is no need to be constrained by our old paper-based technologies

but rather we should be building the best possible assessments to assess what we really want students to know and be able to do. This will not only provide more relevant and better information to guide their futures but will also encourage teachers to implement more appropriate pedagogies. For high-stakes assessment there is no longer the need to sacrifice construct validity to manageability and reliability (Stobart, 2008), we can aim to satisfy all three constraints.

Although we have not yet realised Taylor's vision, as a result of research over the past seven years we are now much closer. It remains to convince some students, many teachers and most of the public that,

> *Technology holds the potential to revolutionize the delivery and assessment of learning outcomes and will result in fundamental changes in how we teach; which mental processes, skills and understandings we measure; and how we make decisions about student learning. (Taylor, 2005, p. 102)*

As digital technology continues to rapidly evolve there is almost no approach to assessment that we couldn't envisage some means of facilitation with the technology. For example, in our research peer collaboration, oral response and modeling were all supported in assessments, and we tested glasses in place of screens to counter 'cheating'. Therefore, in assessing we should focus on what is valuable to assess, and what is considered to be the most valid means of assessment, and then customise the technology to support manageable implementation and reliable measurement. No longer is the technology the limitation, it is our collective imagination and will that will require the fostering of a shared understanding of this focus within our community (Lesgold, 2009).

Acknowledgments

References

A. Willock Information Systems. (2013). *Web-based testing*. Retrieved January 23, 2013, from http://www.awinfosys.com/main/public/website.asp

Becta. (2010). *Messages from the evidence: Assessment using technology*. Retrieved January 14, 2013, from http://labspace.open.ac.uk/file.php/8266/mfte_assessment.pdf

Bennett, J., Tognolini, J., & Pickering, S. (2012). Establishing and applying performance standards for curriculum-based examinations. *Assessment in Education: Principles, Policy & Practice, 19*(3), 321–339.

Bennett, R. E. (2002). Inexorable and inevitable: The continuing story of technology and assessment. *Journal of Technology, Learning, and Assessment, 1*(1), 1–23.

Boyle, A. (2006). *An evaluation of the decision to base the key stage 3 ICT test on a bespoke virtual desktop environment*. London, U.K.: Qualifications and Curriculum Authority.

British Broadcasting Corporation. (2009). Norway tests laptop exam scheme. *BBC News*. Retrieved May 1, 2009, from http://news.bbc.co.uk/2/hi/technology/8027300.stm

Bull, J., & Sharp, D. (2000). *Developments in computer-assisted assessment in UK higher education*. In R. Sims, M. O'Reilly & S. Sawkins (Eds.), conference proceedings, *Learning to choose: Choosing to learn* (pp. 255–260), Queensland, Australia: ASCILITE.

Cisco, Intel, & Microsoft. (2009). *Transforming education: Assessing and teaching 21st century skills*. Retrieved June 8, 2009, from http://atc21s.org/wp.../Cisco-Intel-Microsoft-Assessment-Call-to-Action.pdf

Clarke-Midura, J., & Dede, C. (2010). Assessment, technology, and change. *Journal of Research on Technology in Education, 42*(3), 309–328.

Dede, C. (2008). Theoretical perspectives influencing the use of Information Technology in teaching and learning. In B. K. J. Voogt (Ed.), *International handbook of information technology in primary and secondary education* (Vol. 20, pp. 43–62). New York: Springer Science + Business Media, LLC.

Dochy, F. (2009). The Edumetric quality of new modes of assessment: Some issues and prospects. In G. Joughin (Ed.), *Assessment, learning and judgement in higher education* (pp. 85–114). Wollongong: University of Wollongong.

Fluck, A., Pullen, D., & Harper, C. (2009). Case study of a computer based examination system. *Australasian Journal of Educational Technology, 25*(4), 509–523.

Harris, S. (2008). *Governing Board awards WestEd $1.86 million contract to develop first-ever technological literacy framework*. Retrieved October 7, 2008, from http://www.nagb.org/newsroom/release/tech-literacy-100608.pdf

Heldsinger, S., & Humphry, S. (2010). Using the method of pairwise comparison to obtain reliable teacher assessments. *The Australian Educational Researcher, 37*(2), 1–20.

Kimbell, R., Wheeler, T., Miller, A., & Pollitt, A. (2007). *e-scape: e-solutions for creative assessment in portfolio environments*. London: Technology Education Research Unit, Goldsmiths College.

Kozma, R. B. (2009). Transforming education: Assessing and teaching 21st century skills. In F. Scheuermann & J. Bojornsson (Eds.), *The transition to computer-based assessment* (pp. 13–23). Ispra, Italy: European Commission. Joint Research Centre.

Lesgold, A. (2009). *Better schools for the 21st Century: What is needed and what will it take to get improvement*. Pittsburgh: University of Pittsburgh.

Llamas-Nistal, M., Fernández-Iglesias, M. J., González-Tato, J., & Mikic-Fonte, F. A. (2013). Blended e-assessment: Migrating classical exams to the digital world. *Computers & Education, 62*, 72–87.

McNulty, J., Chandrasekhar, A., Hoyt, A., Gruener, G., Espiritu, B., & Prince, R. (2011). Computer-based testing in the medical curriuclum: A decade of experiences at one school. *Journal of Educational Computing Research, 45*(3), 287–295.

Ministerial Council on Education, Employment, Training and Youth Affairs. (2007). *National Assessment Program – ICT Literacy Years 6 & 10 Report 2005*. Carlton South, Australia: Curriculum Corporation.

National Center for Education Statistics. (2012). The Nation's Reports Card: Writing 2011 (NCES 2012-470). *National Assessment of Educational Progress at Grades 8 and 12*. Washington, DC: National Center for Education Statistics, U.S. Department of Education.

Newhouse, C. P. (2006). Assessment of students by a teacher with a hand held device and a networkable database. In B. L. Mann (Ed.), *Selected styles in web-based educational research* (pp. 220–233). Hershey, PA: Information Science Publishing.

Newhouse, C. P. (2011). Comparative pairs marking supports authentic assessment of practical performance within constructivist learning environments. In R. F. Cavanagh & R. F. Waugh (Eds.), *Applications of rasch measurement in learning environments research* (pp. 141–180). Rotterdam, The Netherlands: Sense Publishers.

Newhouse, C. P. (2012). Digital forms of assessment: Aligning with pedagogic and curriculum intentions. In C. P. Newhouse & J. Pagram (Eds.), conference proceedings, *Australian Computers in Education Conference* (pp. 1–7). Perth, Western Australia: Australian Council for Computers in Education.

Newhouse, C. P. (2013). Computer-based practical exams in an applied Information Technology course. *Journal of Research on Technology in Education, 45*(3), 263–286.

Newhouse, C. P., & Cooper, M. (2013). Computer-based oral exams in Italian language studies. *ReCALL, 25*(3), 321–339.

Newhouse, C. P., & Njiru, J. (2009). Using digital technologies and contemporary psychometrics in the assessment of performance on complex practical tasks. *Technology, Pedagogy and Education, 18*(2), 221–234.

NSW Department of Education and Training. (2013). *Essential Secondary Science Assessment (ESSA)*. Retrieved January 23, 2013, from http://www.schools.nsw.edu.au/learning/7-12assessments/essa/

Pagram, J., & Williams, P. J. (2012). Do you want authenticity with that? Online practical exams in an Engineering course. In C. P. Newhouse & J. Pagram (Eds.), conference proceedings, *Australian Computers in Education Conference*. Perth, Western Australia: Australian Council of Computers in Education.

Panizzon, D., Elliott, K., & Westwell, M. (2010). *e-assessment pilot project. Adelaide: Flinders Centre for Science Educaiton in the 21st Century*. Retrieved December 21, 2012, from www.flinders.edu.au/science...21...2/e-assessment_Feb2010.pdf

Pearson Education Australia. (2012). *Versant*. Retrieved July 12, 2012, from http://www.versanttest.com

Pellegrino, J. W., & Quellmalz, E. S. (2011). Perspectives on the integration of technology and assessment. *Journal of Research on Technology in Education, 43*(2), 119–134.

Penney, D., Jones, A., Newhouse, C. P., & Campbell, A. (2012). Developing a digital assessment in senior secondary Physical Education. *Physical Education and Sport Pedagogy, 17*(4), 383–410.

Ripley, M. (2009). Transformational computer-based testing. In F. Scheuermann & J. Bojornsson (Eds.), *The transition to computer-based assessment* (pp. 92–98). Ispra, Italy: European Commission. Joint Research Centre.

Scheuermann, F., & Bojornsson, J. (Eds.). (2009). *The transition to computer-based assessment*. Ispra, Italy: European Commission. Joint Research Centre.

Stobart, G. (2008). *Testing times, the uses and abuses of assessment.* Abingdon: Routledge.

Stobart, G. (2010). *Assessment fit-for-the-future*. Paper presented at the International Association for Educational Assessment, Bangkok, Thailand. http://www.iaea.info/documents/paper_4d520c71.pdf

Taylor, A. R. (2005). *A future in the process of arrival: Using computer technologies for the assessment of student learning*. Kelowna, British Columbia: Society for the Advancement of Excellence in Education.

Way, W. D., Davis, L. L., & Straine-Seymour, E. (2008). The validity case for assessing direct writing by computer. *Pearson Assessments and Information White Papers*. Iowa City, IA: Pearson.

Weir, A., & Connor, S. (2009). The use of digital video in physical education. *Technology, Pedagogy and Education, 18*(2), 155–171.