

Williams, P.J. (2012). Maximising the potential of ICT to provide authentic summative assessment opportunities. *Computers in New Zealand Schools*, Vol. 24, No. 2, pp. 137–155.

Maximising the potential of ICT to provide authentic summative assessment opportunities

P. John Williams

jwilliam@waikato.ac.nz

Centre for Science and Technology Education Research (CSTER)

The University of Waikato, Hamilton, New Zealand.

Abstract

This paper reports on elements of a study that was conducted in Western Australia to explore the potential of various forms of digitally based external assessments for senior secondary school courses. One problem that needed addressing was how to provide students with authentic assessment opportunities, particularly in subjects in which performance is an integral component. Traditionally, assessment in many of these subjects was by way of a three-hour paper examination. This established a dichotomy for teachers in which the pedagogy of the subject was very different from the method of assessment. In wanting to maximise their student's potential for success, many teachers taught to the examination, consequently sacrificing a practical performance approach to the subject for a more theoretical form of delivery.

Introduction

Students are motivated by the performance aspects of many practical courses such as designing and developing a product, undertaking a conversation in Italian or resolving specific tactical problems encountered in a given sporting context. But because teachers are accountable, they will, as President Barak Obama puts it, tend to “teach to the test”

(eSchool News, 2011, p. 15; Lane, 2004; Ridgway, McCusker & Pead, 2006). Further, educators are also accountable for the outcomes of the use of resources in education (Kleinhenz & Ingvarson, 2004), and there is an expectation that students should demonstrate practical performance not just theoretical knowledge. Moreover, students are more likely to experience deep learning through complex performance (Lane, 2004). As McGaw explains, this places a responsibility on education authorities to consider strategies to increase the assessment of performance on practical tasks.

If tests designed to measure key learning in schools ignore some key areas because they are harder to measure and attention to those areas by teachers and schools is then reduced, then those responsible for the tests bear some responsibility for that. (McGaw, 2006, p. 3)

Performance-based assessment is not new (Messick, 1994). Oral and laboratory examinations have been used in European schools and universities for over a century. In many industries performance-based assessment approaches are used (e.g. for pilot training). In many high-stakes courses in developed countries performance is, and has been, assessed using observation, interview, portfolio or audiovisual recording. In many educational jurisdictions there has been a history of performance-based assessment in some courses such as the arts. However, the use of performance-based assessment in high-stakes courses has been limited by the costs involved in collecting the evidence of performance and difficulties in ensuring reliable and valid results.

Recent advances in psychometric methods and improvements in digital technologies provide tools to assess a variety of performance results relatively cost-effectively. However, any approach or strategy will not be perfect and will require compromises. Many courses in senior secondary schools have some practical components and there is an expectation from students and community that the assessment of student performance will reflect the nature of this learning. In Western Australia, the context for this research, students are externally assessed in most courses using traditional paper-based examinations. In most cases it is clear that performance on practical tasks cannot be assessed adequately on paper. Many educational researchers argue that traditional

assessment only measures knowledge of basic facts and procedures (Lin & Dwyer, 2006), but fails to assess learning processes and higher-order thinking (decision-making, reflection, reasoning and problem solving).

Method

This research was evaluative within an ethnographic framework, in that the activity occurred within learning environments where the culture created by the teachers and students is critical to an understanding of curriculum and pedagogy, including assessment. The study drew on the traditions of interpretive research but also employed, where appropriate, the quantitative methods of more traditional positivist research. That is, quantitative measures concerning achievement, costs and resource use were used but need to be interpreted within the context of the learning environment in which the measures occurred.

The main research question was:

How are digitally based representations of student work output on authentic tasks most effectively used to support summative assessments of student performances for courses with a substantial practical component?

The study was conducted in three phases each of one-year duration. Each phase of the study involved an action research cycle comprising: exploration of the alternatives, development of the prototypes, implementation and evaluation of the prototypes. The three phases were:

- **Proof of concept:** The aim was to explore the feasibility of particular digitally based formats for external assessment in which a wide range of technologies were trialled in order to determine their appropriateness and success within the digital assessment context.

- **Prototype:** In the second year the focus was to develop a full and robust prototype of the system, in which the tasks were refined and the technology streamlined in order to determine the issues of scalability. The aim was to implement a prototype of the form of assessment under ‘normal’ or typical conditions. This included the use of direct capture of student work into the online repository.
- **Scalable product:** In the third year scalability issues were addressed further by limiting deployment to those technologies already available within the trial schools (web-based) and with a USB solution for those schools with poor internet speeds or other restrictions.

The focus of the research was on the use of digital technologies to ‘capture’ performance on practical tasks for the purpose of high-stakes summative assessment. The purpose was to explore this potential so that such performances could be included to a greater extent in the assessment of senior secondary courses, in order to increase the authenticity of the assessment in these courses. The research involved case studies for the four courses involved. During the three years a total of 81 teachers and 1015 students were involved. The number of students involved in each case study ranged from two to 45. Therefore, caution needs to be taken in interpreting the analysis and generalising from the results.

Four different fundamental forms of assessment (reflective portfolios, extended production exams, performance tasks exams, and oral presentations) were investigated in 81 cases with students from the four courses and with the assessment task being different in each course. For each course there was a common assessment task that consisted of a number of sub-tasks. For each case a variety of quantitative and qualitative data was collected from the students and teachers involved, including digital representations of the students’ work on the assessment tasks, questionnaires and interviews. Questionnaires were employed to collect data from students on their characteristics, including perceived level of ICT skills and experience, and their experience of the assessment task. These data were used to address the manageability

dimension of Kimble and Wheeler's (2005) feasibility framework. The questionnaires were administered to all students from the sample classes as soon after the assessment tasks as possible.

Interviews were used to elicit the experiences of students in completing assessment tasks, of teachers in supporting them in these, and of assessors in marking the student work to address the manageability, technical, functional and pedagogic dimensions of the feasibility framework. A stratified sample of students in each class was interviewed as a group, and all teachers, as soon after completion of the assessment tasks as possible. These data were analysed and used to address the research question within a feasibility framework (Kimbell and Wheeler, 2005) consisting of four dimensions:

- *Manageability*: Can the assessment task be reasonably managed in a typical school?
- *Technical*: Can existing technologies be adapted for assessment purposes?
- *Functional*: Can the assessment be marked reliably and validly when compared to traditional forms of assessment?
- *Pedagogic*: Does a digital form of assessment support and enrich students' learning experiences?

The evidence of performance generated from the digital assessment tasks were marked independently by two external assessors using an analytical standards-referenced method. This method used detailed sets of criteria, represented as rubrics, and linked to the assessment task, appropriate course content and outcomes. Correlations were determined for comparison purposes between the two external assessors and also between the assessors and the classroom teacher. Additionally, the work was marked using the method of comparative pairs (Kimbell, et al. 2007; Pollitt, 2004) and these results were again compared against the results from the other forms of marking. Comparative pairs assessment involved a panel of between 5 and 20 assessors and is

based on Rasch modelling (Kimbell, et al. 2007). Each assessor is presented with work from two students, side by side on a computer screen, and makes a judgement about which is better, based on developed holistic criteria. When the group of assessors make enough judgements, a very reliable (for example .90 reliability coefficient for the engineering portfolio assessment) rank order of students is generated.

Findings

The study involved senior secondary classes in four courses: applied information technology, engineering studies, Italian studies, and physical education studies. These were selected as being representative of a broader range of subjects that had similar characteristics to which the findings could be generalised. Following is a summary of the findings of the research for each course.

Applied information technology (AIT)

The intent of the research in AIT was that as students should use computer technology to address challenges in the course, and the summative assessment should align with this intent. At the time summative assessment in the course only involved a three-hour handwritten exam. In general terms, the performance to capture in AIT was the student's response to a challenge given in the form of a design brief to design and development a prototype digital product. Thus the end product is necessarily captured in digital form. However, the design and development processes also needed to be captured digitally with the student creating them in digital form or filming, photographing or scanning. The study considered two types of summative assessment, a digital portfolio and a computer-based exam, with the main differences being the limited time, challenge and set of processes possible in the computer-based exam. The study found that for both the portfolio and the exam it was possible in normal schools to digitally capture a substantial and critical component of student performance. Both forms of assessment allowed for a range of levels of performance.

Each year the production/performance exam had a different design brief but had a similar structure. In response to recommendations by the teachers involved, the design brief was made increasingly more open-ended and provided increasing choices of product types. Each year teachers and students acknowledged the assessment tasks to be faithful to the course syllabus. Completion of the portfolio was facilitated by the teacher, while the exam was facilitated by a researcher and the teacher.

Students in all classes tended to indicate that the digital portfolio was familiar to them although the strict structure was less familiar. Generally, students indicated a preference for the assessment of practical performance at a computer, whether by portfolio or exam. Almost all preferred practical assessment to paper-based theory exams, and believed that they were better able to show their capability using a computer.

All the teachers were positive about the use of the portfolio and exam to assess student performance.

They generally believed that both worked well although they were most comfortable with the portfolio. While they liked the flexibility of the portfolio they recognised the greater ease of invigilating the exam. At least three of the teachers referred to limitations of their infrastructure, including difficulties in working with audio, slow workstations, problems with servers and networks, upgrades to systems and software, and some generally unreliable hardware. However, these concerns largely related to the portfolio and were not evident during the exam.

There was a fine balance between providing adequate scaffolding and allowing students to demonstrate understanding by making choices within open-ended situations. When students were permitted greater choice and freedom that appeared to allow higher achieving students to demonstrate their capability. Greater structure did assist the marking process and therefore there may be a danger of unnecessarily stifling the opportunity for students to demonstrate understanding through over-structuring tasks for marking convenience.

An important finding of the study was that in the third year for all cases the required IT infrastructure was adequate enough to ensure the assessment task could be completed to an acceptable level. This involved a much larger and more diverse sample than for the previous two years. It was found that, provided files were given to students in a limited set of formats, all students could access the required material. It was found to be most efficient to either limit file formats to HTML, PDF, JPG, GIF or WMV, or otherwise convert files to these formats and set limits to the size of media files.

Engineering Studies

Over the three years of the research project there were 11 schools, 14 teachers and 21 classes of Years 9, 11 or 12 students involved in attempting an engineering studies production exam. A case study was constructed for each school. The focus of the digital assessment task was on the core of the engineering studies course which covers:

- engineering design process,
- common engineering principles, structures and systems, and
- enterprise, environment and community.

The assessment was a scaffolded series of sub-tasks developing from a problem scenario to a final solution through a series of design iterations starting with a design brief. Each iteration of the design was reflected in a sketch of a revised idea by the student following some form of stimulus input, which provided the student with information relevant to the solution. The design development culminated in the creation of a 3D physical model and students had three hours to complete all the sub-tasks. A digital design portfolio emerged through the input of a range of forms of data – text, graphics, photographs, voice and video. In the first year a custom-built *Filemaker Pro* database system was used to present the task and collect the digital output, while in the final two years the *e-scape* exam management system (Kimbell et al. 2007) was used to design and present the task.

In the final two years the task was managed in three different ways across the schools. The first of these was via the school intranet using ASUS EeePC mini computers that were wirelessly linked with the facilitator's computer. The facilitator was then able to monitor each logged-in student's progress throughout the examination tasks, and progress all the students on to the next task at the same time. The students' examination portfolios were automatically uploaded to the facilitator's computer at the conclusion of the exam. They then had to be later uploaded to the server in preparation for assessment. The second method was live via the internet and the students worked on school computers and logged on live to the examination server. The research facilitator was also logged on as the exam manager and was then able to monitor each logged-in students progress throughout the sub-tasks, and move the students on when the time expired. The students' exam portfolios were automatically and progressively uploaded to the server. The third method involved utilising a USB, and in this scenario the exam was accessed by the students through logging on to the software on a USB drive inserted into a school computer. This did not enable centralised control by the facilitator but students progressed through the exam at their own rate. The USBs were collected at the end and later uploaded to the server for access by the assessors. Although these different technologies were used for the exam, depending on whether the intranet, live, or USB process was used, the appearance was the same for all students.

Overall, with respect to the assessment task, students indicated that using a computer made all of the sub-tasks easier and that they preferred practical work to a focus on theory. Generally they indicated that they believed that the exam did provide an adequate opportunity to demonstrate their skills and understanding, although some students, and also some teachers, felt that the task was not aligned with what they considered was an engineering problem.

The collation of student work in digital form had several obvious advantages for marking, such as ease of storage, backup, transmission, access and sharing. The web-based database, which held the student work, was responsive and easy to use provided that adequate internet bandwidth was available to the assessor. With regard to the analytic marking, the ability to view both the work sample with the marking rubric

alongside it was convenient and ensured focus was maintained. Switching rapidly between different aspects of the student portfolio was easy using a series of linked buttons named according to the sub-task. The database recorded and summed the scores and this was obviously quick and accurate. Also in each year of the project comparative-pairs holistic judging of student results was undertaken in addition to the analytical marking.

Overall, the three years of study demonstrated that a computer-based production exam could be constructed for the core content of the engineering studies course and that this could be readily implemented in a large range of schools offering the course. The structure of the exam was perceived by some to be over-scaffolded with four iterations of design solutions required in response to stimulus material. This could be overcome with a more comprehensive introduction to the examination and practice by way of mock exams prior to sitting the exam. The time for exam appeared to be adequate, as indicated by the students who paced themselves through the exam using the USB drives. The 15 minutes practice session to familiarise the students with the use of the digital equipment proved to be a necessary element.

Italian studies

The main focus over the three years of this research in Italian was on oral communication. The current external assessment system assesses oral communication by having students travel to a central location and undertake an interview with two external assessors that judge student performance in real time during the interview. This method involves logistical difficulties for both students and the organising body. Additionally, the real-time assessment and lack of an enduring record of the process raises questions about the reliability of this form of assessment. Over the course of this project a number of approaches were used to assess oral performance. Also in the third year other outcomes of listening and responding; viewing, reading and responding were also assessed.

The first year of the study involved four schools with students in Year 11 and the assessment task consisted of a portfolio, which was made up of a number of sub-tasks leading up to a video-recorded oral presentation. It was concluded that the technology available was adequate for all schools to complete the tasks. The oral presentation was video-recorded for students from all the sample schools. Constraints included the time-consuming nature of the portfolio tasks that the students had to undertake, and the nervousness of the students about being video recorded. The main benefits were that the students found that the digital technologies enabled them to reflect critically on their practical performance.

In the second year of the study, students completed the assessment task that consisted of an in-class computer-based exam (a series of sub-tasks completed in a number of separate sessions) and a recorded interview. Teachers and students felt that the sub-tasks in the exam were of an appropriate standard and that the stimuli were well selected. It was not possible to carry out the tasks 'live' on the internet, which meant there was an onerous and time-consuming task in preparing the USB drives, and subsequently uploading each student's work to the server. It was concluded that this method would not be scalable to a large number of schools and the firewall issues with the internet would need to be solved.

The teachers believed that the tasks did not accurately reflect a conversation and would be better if modified so that the students listened to, rather than read, the stimulus questions. In general, the teachers did not believe that a computer-based task could simulate a conversation. However, both the teachers and students felt the tasks were useful preparation for the recorded interview. There were no technical issues with implementing the recorded interview exam, which included two assessors, the student, and a digital video camera in the room.

In the third year the formal part of the assessment task involved Year 11 students completing a computer-based online exam with two components: listening and responding, and oral communication. In the first component students listened to an audio programme in Italian and answered questions, and in the second component

students watched three video segments and made audio recordings in Italian in response to the videos. The assessment task as a whole was acknowledged to be faithful to the course and it successfully distinguished between students with various levels of mastery of the Italian language.

Both students and teachers found the listening and responding task to be equal to, or better than, the traditional form of assessment of this outcome. However, the oral communication component was considered by both teachers and students to be inferior to the traditional face-to-face oral exam. This was due to both technical problems and the nature of simulating a conversation.

In the third year of the study all student work was stored in digital form to be accessed online by assessors. The assessors reported that both the analytic and comparative methods of marking were easy to use. Overall the third year of the study found that the benefits of the digital forms of assessment implemented outweighed the constraints. The major benefit was that students could proceed at their own pace; and having their own set of headphones, had a distraction-free listening experience. In general schools have the required technologies to undertake digital assessments in the course that deliver media (text, audio, and video) and capture student performance, whether that is in the form of text, audio, or video-based.

Physical education studies (PES)

The aim of the research for the PES course was to enhance the practical examination that focuses on students' ability to make decisions and apply skills to resolve tactical problems encountered in a sporting context. The digital assessment aimed to address issues of sustainability and authenticity while also ensuring rigour in an examination context. The concern for sustainability reflected the growing numbers of students taking PES and students being geographically dispersed. In relation to rigour and authenticity, the aim was to integrate some of the theoretical content of the course in the four components of the performance exam by capturing the intentions of students in their strategic responses, requiring them to adapt responses to changing circumstances,

asking them to apply theoretical knowledge, concepts and principles to a given situation, and reflect on or evaluate their performance. The task was designed to be adapted to contrasting sporting contexts used in PES.

Over the three years, there were 11 schools, 18 teachers and 19 classes of Year 11 students involved in the research associated with the PES course. In the first year of the study, the assessment task was implemented at three schools with a total of four classes of Year 11 students. Each of the four classes involved a different sport and thus there were four sporting contexts: rugby union, volleyball, swimming and soccer. The following year, the assessment task was implemented at two schools with a total of four classes of Year 11 students in four sporting contexts: cricket batting, cricket bowling, soccer and swimming. In the third year of the study, the assessment task was implemented at six schools with a total of eleven classes of Year 11 or 12 students in seven sporting contexts: cricket batting, cricket bowling, soccer, swimming, netball, volleyball and tennis.

A remote-controlled four-camera system was used to capture student performance. The placement of the cameras varied for each sport. Video feeds from the individual cameras were able to be displayed as quarter screen or full screen images for use in student critique of their own performance and in assessment. The underwater camera used to capture swimming performance was custom designed. *Quicktime Pro* software was used to crop, cut and paste each student's recordings into separate files that were then converted to WMV files by batch using appropriate software

All four components of the performance task exam were completed successfully with all groups of students with varying degrees of technical and logistical obstacles to overcome. Generally, students regarded the task as an appropriate means of assessment for the course and perceived it to be authentic and meaningful and preferred to a written exam. The task effectively encompassed conceptual, practical and reflective aspects that were able to be adapted for application and implementation in varied sporting contexts. It was also acknowledged as not dissimilar to the learning experiences that teachers were endeavouring to build into their teaching in the course. In almost all cases the

assessment was perceived to be more authentic than a paper-based exam. The task was considered to have achieved a degree of connection between traditionally distinct 'theoretical' and 'practical' components of PES. The teachers were particularly enthusiastic, indicating that the task was superior to the current practical exam conducted in the course.

The main difficulties encountered by the assessors with both analytical and pairs marking were technical and largely concerning access to the videos. Because the videos were not in a streaming format they had to wait for a video to download before viewing; this increased the time to mark. Also some assessors found particular videos difficult to see (e.g. dark images on overcast days) and had difficulty identifying the student being assessed. Nevertheless, it is notable that marking was conducted effectively by assessors using an online medium and in a manner that allowed for input from assessors located throughout the country. The server-based comparative pairs interface was trialled very effectively and was well received by assessors, except for the degradation of video quality and the observation that it would be preferable to be able to view two student videos 'in tandem' via a split screen, rather than having to switch between students.

The time allocations for the exam appear to be adequate if technical checks are done on site in advance, and if students are also introduced to the computer interface for the task prior to undertaking the exam. Implementation of the exam should initially continue to use USB flash drives for delivery and collection of student work with an online upload at the end of the exam as a backup.

Conclusions

For each course the intention to implement the same assessment task for each case was achieved by the third year for all four courses where, in each case, implementation of the assessment task was facilitated by either a researcher or trained invigilator with assistance from the teacher. The study used a four-dimensional feasibility framework (Kimbell & Wheeler, 2005) to investigate the effectiveness of each form of assessment. Conclusions are summarised below under the headings for the feasibility framework.

Manageability dimension

Overall it was concluded that:

- There were few logistical difficulties for production or performance tasks exams except where longer sessions had to be organised within normal lesson timetables. In non-computing courses booking of computer laboratories may be a logistical difficulty, particularly if a portfolio or series of online tasks are undertaken.
- There were some logistical difficulties in managing portfolios. This was usually concerning the management of student time to ensure they completed all requirements.
- Where assessment tasks were video recorded outside or in a large space such as a gymnasium, it was best if the space was confined; the amount of ambient light even; and using smaller numbers of students with bright identifying apparel was advantageous for recognition of individual students.
- Spacious rooms with at least 10% excess computers have the advantage of allowing students to relocate in the case of technical problems and reduced opportunity to see other students' work.
- Practice sessions were needed to familiarise students with the assessment tasks.
- Processing of locally collected data to be uploaded into an online repository to be accessed with a marking system required resources and management.
- The use of the 'pairs engine' improved the manageability of comparative pairs marking.

Technical Dimension

Overall it was concluded that:

- In almost all cases reliance on school computing infrastructure was enough to successfully implement a digital assessment task. This was less the case if internet access was required.
- Laptops and netbooks were found to be as good as desktop computers, which increases the flexibility of where the exam might be held.

- For online assessment tasks, extensive testing of networks (especially under load) and workstations were necessary.
- Audio- or video-recorded responses to assessment task items need to be stored locally (in addition to any attempted streaming of responses to a server) to minimise data loss.
- Any online examination system needs to be adaptable to a variety of technical environments of hardware, browsers, and operating systems within schools.
- Technical issues are most likely to arise when students are required to make video or audio recordings on a computer using peripherals such as web-cams and/or headsets. Additionally, uploading audio and video-based responses puts the greatest strain on school networks including bandwidth (watching videos or listening to audio, that is downloading, was not a problem).
- The multi-camera remote control video-capture system used in PES proved to be adequately reliable and flexible including use in a range of indoor and outdoor sporting facilities, support for underwater filming for swimming, provision of feedback to facilitator, and potential for control by a single operator. The use of an iPad to observe what was being video-captured was beneficial.

Functional dimension

Overall it was concluded that:

- For all the courses the students and teachers readily perceived the assessment task(s) to be authentic, meaningful and contributed to connecting the theoretical and practical components of the courses. Generally they preferred this to the alternative of a paper-based exam although in Italian the majority preferred face-to-face oral work rather than using a computer.
- For all four courses the assessment task was structured permitting a good range of levels of achievement to be demonstrated. This was reflected in the wide range of scores from marking and the perceptions of teachers and students.
- Digitally based assessment was shown to enable varied forms of student responses (e.g. written, drawing, audio, video) and this was appreciated by students and teachers.

- The comparative pairs method of marking was successfully implemented for all four courses with resulting highly reliable scores. In AIT it was only applied to the performance exam and in Italian only for the recorded oral presentation/interview in the first two years.
- Online assessment of digital artefacts is an effective way to facilitate assessment and offers additional affordances such as access from a variety of locations, efficient storage and backup, and sharing of data over the traditional method.

Pedagogic dimension

Overall it was concluded that:

- Typically, students liked doing the practical work involved in the digital assessment tasks and, apart from in Italian, they preferred the digital form of assessment to paper-based theoretical work. They were happy to answer questions where they could type and draw responses.
- In general, the assessment matched general pedagogy for the course for most classes involved and was viewed positively by most teachers.
- In all four courses the quality of work was highly dependent on the class, probably reflecting differences in capability of the students and pedagogical approaches by the teachers.
- Many students appreciated the opportunity to demonstrate their creative capability in examinations situations.
- Digital assessment provides the ability to capture student knowledge and performance using a number of media (text, images, sound, video, etc.) and this provides an improved and more authentic method compared with the traditional paper-and-pen method of assessment.

Constraints and benefits

Across the four courses the main constraints were logistical in organising time to complete the tasks, and in some cases technical in either running software on school workstations or accessing online systems through school networks. In a number of

schools changes had to be made to standard operating systems to allow software to run off USB drives, videos to be viewed, Flash applications to run within internet browsers and sound to be recorded. There were some difficulties organising adequate amounts of time to complete tasks. For engineering and Italian there were some difficulties negotiating school firewalls to access online assessment systems.

There were different benefits across the three general forms of digital assessment. Reflective portfolios generally permitted students to address a greater range of outcomes and demonstrate a greater range of knowledge and skills although this was not realised in Italian. Production exams generally permitted students to address a limited range of outcomes and demonstrate a reasonable depth of knowledge and skills dependent on the task and time. Performance task(s) exams generally permitted students to address a range of outcomes and demonstrate a limited depth of knowledge but a reasonable level of skill dependent on the task and time. For AIT and engineering, both students and teachers were conspicuous in their support for the use of portfolios and computer-based exams in place of paper-based theoretical exams.

The research found that the benefits outweighed the constraints. In particular, student responses tended to be overwhelmingly positive, typically due to the practical nature of the work and relevance to their interests. Generally they preferred this to paper-based exams. In Italian, although students were not as positive, they did indicate that it was valuable being able to critically reflect on their own performance by viewing the videos.

References

- eSchool News. (2011). Obama says too much testing makes education boring. *eSchool News*, 14(5), 15.
- Kimbell, R., & Wheeler, T. (2005). *Project e-scape: Phase 1 report*. London: Technology Education Research Unit, Goldsmiths College.

- 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.
- Kleinhenz, E. & Ingvarson, L. (2004). Teacher accountability in Australia: Current policies and practices and their relation to the improvement of teaching and learning. *Research Papers in Education*, 19(1), 31–49.
- Lane, S. (2004). Validity of high-stakes assessment: Are students engaged in complex thinking? *Educational Measurement, Issues and Practice*, 23(3), 6–14.
- Lin, H., & Dwyer, F. (2006). The fingertip effects of computer-based assessment in education. *TechTrends*, 50(6), 27–31.
- McGaw, B. (2006). *Assessment to fit for purpose*. Paper presented at the 32nd Annual Conference of the International Association for Educational Assessment, Singapore.
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23(2), 13–23.
- Pollitt, A. (2004, June). *Let's stop marking exams*. Paper presented at the International Association for Educational Assessment Conference, Philadelphia, USA.
- Ridgway, J. McCusker, S. & Pead, D. (2006). *Report 10: Literature review of e-assessment*. UK: Futurelab.



This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivs 3.0 New Zealand License](https://creativecommons.org/licenses/by-nc-nd/3.0/nz/).