1-1-2012

Digital forms of assessment: aligning with pedagogic and curriculum intentions

Christopher Newhouse

DIGITAL FORMS OF ASSESSMENT: ALIGNING WITH PEDAGOGIC AND CURRICULUM INTENTIONS

C. Paul NEWHOUSE
Associate Professor, School of Education, Edith Cowan University

Abstract

Increasingly in the world fewer work tasks are done using paper and pen and yet most high-stakes assessment in schools continues to use this primitive technology. In the past it has been considered too difficult to reliably and manageably assess large cohorts of students using approaches more valid than using paper. The range of maturing digital technologies for handling multimedia now provides opportunities to address this discrepancy. This paper reports on a three-year project investigating the use of digital technologies to represent student work for high-stakes summative assessment in four senior secondary courses. The project used a range of digital devices, online and offline technologies, and database portfolio systems linked to online marking tools. In addition the comparative pairs method of marking was compared with an analytical standards-referenced method. The project involved 82 teachers and their senior secondary classes.

Introduction

Over a decade ago a review of senior secondary schooling in Western Australia (W.A.) led to a range of new courses such as *Applied Information Technology* (AIT), that were based on subjects previously designated as non-tertiary entrance and others, such as *Physical Education Studies* and *Engineering Studies*, that represent new areas of study, and enhancements to existing courses, such as *Italian Studies*. In many of the courses there were major practical components and there was an expectation that the assessment would reflect these. However, for most courses students are externally assessed using paper and pen that is clearly inadequate. Many leading educators (Lane, 2004) argue that this traditional approach to assessment is inadequate to assess higher-order thinking. In the world fewer work tasks are done using paper and pen. Lin and Dwyer (2006) suggest that digital technologies need to be used to capture "more complex performances" (p. 29) that assess a learner's higher-order skills but that due to "technical complexity and logistical problems" (p.28) this has been slow to materialize. Thus there is a critical need for research into the use of digital forms of assessment on complex tasks that are feasible within schools.

What is assessed is critical because students tend to focus on, and be motivated by these sections of the curriculum, and teachers, as President Barak Obama puts it, tend to “teach to the test” and make “education boring for kids” (eSchool News, 2011). Further, society is increasingly expecting that students should demonstrate practical performance not just theoretical understanding. Finally, students are more likely to experience deep learning through complex performance. Therefore as McGaw (2006) explains, this places a responsibility on education authorities to consider performance assessment strategies.

In W.A. there has been a history of performance-based assessment in some courses in the Arts but this has been limited by the costs involved and difficulties in ensuring reliable and valid results. However, recent advances in psychometric methods and digital technologies provide tools to assess a variety of performance relatively cost-effectively (McGaw, 2006). This was demonstrated by the British e-scape project (Kimbell, Wheeler, Miller, & Pollitt, 2007) with student work represented in digital form, collated using an online repository and marked using a comparative pairs judgements method.

The Project

This paper reports on a three-year project conducted by the Centre for Schooling and Learning Technologies (CSaLT) at Edith Cowan University (ECU) in collaboration with the W.A. Curriculum Council. The focus of the study was on the use of digital technologies to ‘capture’ performance on

practical tasks for the purpose of high stakes summative assessment. This may include representations of student 'outputs' or performance processes. The purpose was to increase the authenticity of the assessment with a focus on four senior secondary courses: *Applied Information Technology* (AIT); *Engineering Studies*; *Italian Studies*; and *Physical Education Studies* (PES). The study included 81 class-based case studies each involving a teacher and a class in one of the four courses. In total there were 1015 students involved. The first year of the study formed a *Proof of Concept* stage followed by a *Prototype* stage and finally, in the third year a *Scalable Product* stage with a representative sample.

At the beginning a research team developed a common assessment task with standardised parameters for each of the four courses. This started with a situation analysis that considered the characteristics of students, the requirements of the course, the nature of the performance to be assessed, the technologies that could be used to capture this performance, and the characteristics of typical teachers. From this the structure of each assessment task was formed and examples sought. During this process teachers were recruited to implement the assessment tasks with their classes.

A framework of four dimensions that was developed for the e-scape research (Kimbell et al., 2007) was used to evaluate feasibility by combining the range of data including observation in class, a student survey, interviews with teachers and groups of students, student work output, and the assessment records of the teacher. Output from student work on the assessment tasks was collected in digital form and placed in the online digital repository to be available to the markers. The work was assessed by two external markers using an analytical 'rubric' method and by a panel of five or more using a comparative pairs method. For each method a set of scores and rankings were generated. These were compared using tests of correlation.

The Assessment Tasks

For *AIT* a hybrid assessment task structure was developed to compare a digital portfolio with a production/performance computer-based exam. However, in the final year only the exam was implemented. The multi-part reflective process portfolio included a digital product, the collation of a process document associated with creating the digital product, and two other digital artefacts. The exam required students to design, produce and evaluate a digital product associated with a design brief challenge. The tasks were completed using USB flash drives on school computer workstations.

For *Engineering Studies* the assessment task was a scaffolded series of sub-tasks developing from a problem scenario (e.g. rudimentary water distillation system) to a final solution through a series of design iterations. Each iteration of the design was reflected in a sketch following some form of stimulus input. In the first two years this three-hour exam culminated in the creation of a physical model, however, this was removed in the final year and the time reduced to two hours. A digital design portfolio emerged through the input of a range of forms of data – text, graphics, photographs, voice and video – using either an online or offline management system. In the first year the task was managed using a custom-built database, while in the final two years the e-scape system was used.

For *Italian Studies* the assessment task evolved considerably over the three years but with a continuing focus on an *oral communication* outcome. In the first year the assessment consisted of a *Portfolio*, made up of a number of sub-tasks, leading up to a video-recorded *Oral Presentation*. In the second year the portfolio was replaced with an in-class computer-based *Exam* (a series of sub-tasks completed in a number of sessions) and a *Recorded Interview*. In the third year these were replaced with a computer-based online *Exam* of two components - listening and responding, and oral communication. In the first students listened to an audio program in Italian and typed answers to questions, and in the second they watched three videos and made audio recordings in Italian in response to the videos.

For *PES* the assessment task was a four-component performance tasks exam including response questions. The first component set a tactical challenge situation for a sport and asked questions to analyse the challenge and propose solutions by typing responses using a database. The second

component required students to undertake four drills to demonstrate skills relevant to the challenge. The third component involved modified 'game' situations associated with the tactical challenge. Performances were video recorded using a remote control multi-PTZ-camera system. The final component required students to view videos of their performances and respond to questions using a database with links to their digital video files.

Data Analysis

The range of types of data collected were initially analysed separately for each class and source of data and then combined for each course and then between courses.

Observations, Interviews and Surveys

All classes were observed completing the components of the assessment task. On completion of the task the students were surveyed, and the teacher was interviewed as well as a representative group of students. Students completed a questionnaire consisting of 57 closed response items and two open-response items. A number of scales were derived from the questionnaire (refer to Table 1).

Table 1
Descriptions and mean (standard deviation for AIT) statistics for the scales based on items from the student questionnaire.

	AIT*	Eng	Ital	PES	Description
N	94	88	86	149	
eAssessE	3.0 (0.5)	2.9	2.9	3.0	Value of ICT to exam (1 to 4).
eAssessP*	3.2 (0.6)				Value of ICT to portfolio (1 to 4).
Apply	2.4 (0.4)	2.3	2.3	2.1	Application of computer use (1 to 3).
Attitude	2.6 (0.3)	2.5	2.5	2.4	Attitude towards using computers (1 to 3).
Confidence	2.8 (0.3)	2.7	2.5	2.5	Confidence in using computers (1 to 3).
Skills	3.3 (0.5)	3.2	2.8	2.8	Self-assessment of ICT skills (1 to 4).
SCUse	79 (69)	59	52	58	Estimated mins/day using computers at school.

*Note: There was little difference between the standard deviations for the four courses. * AIT portfolio was not implemented in the third year and therefore the mean and standard deviations quoted are for the second year.*

In general it was found that the AIT, PES and Engineering teachers and students were positive about the assessment task and that its implementation was faithful to the course, and they perceived few difficulties. The Italian teachers and their students tended not to be as positive towards the recording of oral performances but did see value in computer-based oral recording as practice and review. Students in all four courses indicated through the survey the value in using ICT to completing aspects of the assessment tasks (eAssess means around 3.0). On average AIT students indicated that ICT use in the portfolio was the most valuable. As would be expected the mean ICT skills score was lower for Italian and PES students as was their confidence and attitude towards using computers, and the amount of time they used computers at school. As would be expected the AIT students indicated using computers considerably more than the other students.

In AIT students indicated a preference for the assessment of practical performance at a computer, whether by portfolio or computer-based exam. Almost all preferred practical assessment to paper-based theory exams, although less so for final year students, and believed that they were better able to show their capability using a computer in preference to paper. A few had concerns about malfunctions of systems and a lack of time during computer-based exams. In Italian Studies teachers and students acknowledged the assessment task to be faithful to the course and it successfully distinguished between students with various levels of mastery of the Italian language. However, the oral communication component was considered 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 PES teachers perceived that the task effectively encompassed conceptual, practical and reflective aspects that were able to be adapted for application and implementation in varied sporting contexts.

Methods of Marking

Two methods of marking were successfully implemented for the assessment tasks in all four courses: an analytical standards-referenced method; and a comparative pairs method. However, for Italian this was only done for recorded oral components and for AIT only for the production/performance tasks exam component. For the analytical method a set of assessment criteria represented in rubrics was developed for each course. For the comparative pairs method one holistic criterion was used for each assessment task. For both methods of marking the process was supported by the use of online digital marking tools and online portfolio systems.

The scores from two analytical assessors, the comparative pairs marking, and provided by the teachers whose classes were involved were compared using correlation analysis. The correlation coefficients for all the courses for the third year are reported in Table 2. The scores from the two assessors using the analytical method of marking were moderately to highly correlated except for in Engineering. There were moderate correlations with the teacher scores but strong correlations between the two methods of marking for all courses except Engineering. The low correlation with teacher scores was due to a combination of teachers using different criteria, most classes being small samples and teachers taking into account background knowledge of the students. The Cronbach's Alpha values from the Rasch analysis (around 0.90) indicated a high reliability. Engineering assessors had more difficulty scoring consistently analytically compared with using the comparative pairs method.

Table 2
Correlations between assessors and methods of marking for the third year.

		α		Correlations Between Assessors or Methods of Marking				
				Analytic Assessors		A & T	A & P	
Course	N	A	P	Scores	Rankings	Scores	Scores	Rankings
AIT	163	0.90	0.90	0.69**	0.64**	0.60**	0.68**	0.68**
Engineering	94	0.89	0.90	0.53**	0.49**	0.64**	0.46**	0.38**
Italian	97	0.91	0.94	0.88**	0.85**	0.48**	0.77**	0.69**
PES	152	NA[a]	0.96	0.61**	0.56**	0.64**	0.73**	0.74**

*** p<0.01 (2-tailed) * p<0.05 (2-tailed). [a] Not able to calculate due to missing data.*

Results

For *AIT* in the first two years both a portfolio and an exam were implemented for 13 classes, however, the extent to which students completed the portfolio varied considerably. The exam was completed by all students with almost no technical difficulties evident apart from the recording of sound for a few classes. Almost all students indicated a preference for the two forms of assessment and that both provided a good assessment of practical performance. They commented on the ease of working on the computer compared to working on paper … correcting errors, speed of writing, amount of writing, speed of action and physical comfort. When pressed, their major concern was malfunctions of systems during exam. The manner in which the portfolio was implemented varied between classes. Generally the product requirements provided adequate scope to demonstrate capability. The process document varied in quality. The exam provided less scope for demonstration of capability. A major weakness of the exam was that a relatively low level set of tasks was required so that students in all classes could attempt them using a typical set of software. In each year a slightly higher level of tasks was set, as a result of allowing students more choice in the type of prototype product.

For *Engineering* a production exam was implemented for 21 classes. Generally the students indicated that using a computer made all of the sub-tasks easier and that the exam did provide an adequate opportunity to demonstrate their skills and understanding. They found the experience engaging and enjoyable, and felt that it more accurately reflected the nature of the engineering course they were studying. The expectations of the teachers varied, some considered that their students achieved well, others who felt that their students should have been able to produce work of a higher standard. In the final two years the task was managed in three different ways using the e-scape system, but the user

interface for each method was the same. The easiest method to implement was live via the Internet using the school's computers and logged on to the system's server. The invigilator was logged on to control and monitor progress through the sub-tasks. If this method was unavailable due to network restrictions then a local wireless intranet could be set up with the facilitator's computer onto which the exam portfolios were stored to be batch uploaded to the e-scape server. A third method used a USB flash drive with client software that stored the portfolios to be uploaded to the server at a later time.

For *Italian* the nature of the assessment task changed between the three years but with the main focus on recording oral performance. In the first two years an oral performance was video recorded while in the second and third years a component involved computer-based oral recording. Generally students indicated a high level of apprehension that may lead to reduced performance with video recorded oral work. They also indicated little experience in doing assessments using digital technologies. However, teachers and students recognized the value of students being able to critically reflect upon their oral performance through the use of digital recordings of their responses to various stimuli. Online assessment appeared to be equally effective compared with traditional methods and offers other benefits such as convenience and access from a variety of locations. The second year of the project demonstrated that externally marked computer-based oral tasks carried out in class time correlated highly with scores from traditional face-to-face recorded interviews.

For *PES* over the three years, the assessment task was implemented with 19 classes each involved in one of seven sporting contexts: rugby union, volleyball, swimming, soccer, cricket batting, cricket bowling, netball, and tennis. 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 teacher perceived the assessment 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.

Conclusion

To address the research questions, the results of the data analysis were interpreted within the four dimensions of a *Feasibility Framework*.

The *Manageability* dimension concerns making a digital form of assessment do-able in typical classrooms with the normal range of students. In general it was possible to implement each of the assessment tasks successfully. 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 provided some logistical difficulty, particularly where a portfolio or series of online tasks were undertaken. There were some logistical difficulties in managing portfolios, usually time to ensure completion of all requirements. There was little evidence that any students were adversely affected by a lack of ICT experience or capability. Overall the use of online marking tools with digital repositories was manageable with only minor difficulties to overcome and a reasonable level of maintenance and management required.

The *Technical* dimension concerns the extent to which existing technologies can be adapted for assessment purposes within course requirements. Overall there were no significant technical difficulties that could not be easily overcome. In almost all cases reliance on school computing infrastructure was adequate. This was less the case if Internet access was required where for online assessment tasks extensive testing of networks (especially under load) and workstations was necessary. Where online exam systems were used they needed to be adaptable to a variety of technical environments across the variety of hardware, browsers, and operating systems. Audio or video recorded responses to assessment task items needed to be stored locally to minimise data loss. The

multi-camera remote control video-capture system proved to be adequately reliable and flexible for a range of indoor and outdoor facilities, including underwater filming, provision of feedback to facilitator, and potential for control by a single operator.

The *Functional* dimension concerned reliability and validity, and the comparability of data with other forms of assessment. In general 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 Studies the majority preferred face-to-face oral work. The tasks permitted a good range of levels of achievement to be demonstrated. External assessors using an analytical method delivered reasonably high levels of reliability except in Engineering. The comparative pairs method of marking was implemented with resulting highly reliable scores ($\alpha \geq 0.90$). This method appeared to be best suited for performance-based tasks that measured more open-ended tasks with a main outcome. Assessment tasks constructed to measure a number of diverse outcomes were best assessed using the analytical approach. The scores from comparative pairs marking were generally moderately correlated with those from analytical marking except in the case of Engineering. Here assessors found it easier to form a holistic judgment than analyse the performance using a large range of criteria, which had large ranges of score points.

The *Pedagogic* dimension concerns the extent to which the use of a digital assessment forms can support and enrich the learning experience of students. In general, for AIT and PES, the assessment matched general pedagogy for the course for most classes involved and was viewed positively by most teachers. This was less the case for Italian and Engineering teachers, although they generally perceived value in including the tasks. 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. Some courses, such as Italian, have a well-established set of pedagogies tied to particular historic and traditional assessment models and these would take time to change. 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 paper and pen method of assessment.

Across the courses the main constraints were logistical in organising time to complete the tasks and in some cases technical in running software on school workstations or accessing online systems through school networks. However, overall the benefits outweighed the constraints. Student responses were overwhelmingly positive due to the practical nature of the work and relevance to their interests. In Italian, although students were not as positive, they did indicate that it was valuable being able to critically reflect on their own performance. Reflective portfolios permitted students to address a greater range of outcomes and demonstrate a greater range of knowledge and skills. Production exams permitted students to address a limited range of outcomes and demonstrate a depth of knowledge and skills. Performance task(s) exams permitted students to address a range of outcomes and demonstrate a limited depth of knowledge.

In all courses it was found that digital technologies could be implemented to enhance to varying degrees the quality of assessment. This was most notable in AIT where it became clear that without the use of the technologies the intention of the course was likely to be lost. In Engineering the authenticity of assessment for a core component of the course was increased while in PES current practices in assessing sporting skills were enhanced and a connection to core theory was added. Finally, in Italian the integration of outcomes within an assessment task was enhanced and a more cost effective alternative to physical interviews was provided. While employers and community leaders call for schools to produce students with 21st Century skills (Kozma, 2009) and deep conceptual understanding the main driver of curriculum and pedagogy, assessment, continues to focus on 19th Century skills and shallow recall of content. This highlights the need to use digital technologies to support alternative approaches to assessment.

Acknowledgement

The theory discussed in this paper and the research upon which it is based are as a result of the work of a research team organised by the Centre for Schooling and Learning Technologies at Edith Cowan University. The team was led by Paul Newhouse and John Williams (University of Waikato) and included researchers Dawn Penney (University of Waikato), 'Chirp' Lim (University of Hong Kong), Jeremy Pagram, Andrew Jones, Mark Hackling, Ron Oliver, Russell Waugh, Martin Cooper and Alistair Campbell, and a number of research assistants.

References

eSchool News. (2011). Obama: Too much testing makes education boring. *eSchool News, 14*(5), 15.

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.

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.