1-1-2011

Digital Assessment in Secondary Schools: A Western Australian Experience Focusing on a Senior Engineering Course

Jeremy Pagram
Edith Cowan University

John Williams
University of Waikato, New Zealand

T. Hirashima et al. (Eds.) (2011). Proceedings of the 19th International Conference on Computers in Education. Chiang Mai, Thailand: Asia-Pacific Society for Computers in Education

DIGITAL ASSESSMENT IN SECONDARY SCHOOLS: A Western Australian Experience Focusing on a Senior Engineering Course

Dr Jeremy PAGRAM[a*] & Dr P John WILLIAMS[b]
[a] *Senior Lecturer, Edith Cowan University, Australia*
[b] *Associate Professor, University of Waikato, New Zealand*
* *j.pagram@ecu.edu.au*

Abstract: This paper reports on the results of a three-year study conducted at the Centre for Schooling and Learning Technologies (CSaLT) at Edith Cowan University in collaboration with the Curriculum Council of Western Australia which concerns the potential to use digital technologies to represent the output from assessment tasks in four senior secondary courses: Applied Information Technology, Engineering Studies, Italian and Physical Education Studies. This paper focuses on Engineering Studies. The general aim of this study is to explore the potential of various digitally-based forms for external assessment for senior secondary courses in terms of manageability, cost, validity and reliability.

Keywords: Digital assessment, Pairs Marking, Computer Managed Learning, High stakes assessment, Online learning, e-learning

Introduction

This paper describes the results a three-year study conducted at the Centre for Schooling and Learning Technologies (CSaLT) at Edith Cowan University in collaboration with the Curriculum Council of Western Australia and supported by an Australian Research Council grant. The study commenced in January 2008, was completed in December 2010, and concerns the potential to use digital technologies to represent the output from assessment tasks in four senior secondary courses: Applied Information Technology, Engineering Studies, Italian and Physical Education Studies. The work parallels and builds upon research undertaken by Kimbell [2] from Goldsmiths College London. This paper will focus on Engineering Studies, with a particular emphasis on the implementation strategies used.

1. Significance and Rationale

From the 1990s, significant developments in computer technology have included the emergence of low-cost, high-powered portable computers, and improvements in the capabilities and operation of computer networks. These technologies have appeared in schools at an escalating rate. During that same period many school systems were moving towards a more standards-based curriculum and investigating methods of efficiently and effectively assessing students from this perspective. Many of the high-stakes senior secondary courses being implemented over the latter half of the decade have a significant performance component, and are not able to be adequately assessed using paper-based methods. Therefore it is important that a range of forms of assessment are considered along with the potential for digital technologies to support these alternative forms.

While students tend to focus on, and be motivated by practical performance in courses, teachers being accountable for student results will tend to 'teach to the test' [3][6]. Educators are increasingly accountable to society for the outcomes of the use of resources in education, and our society increasingly expects that students should demonstrate practical performance and not just theoretical knowledge. Moreover, students are more likely to experience deep learning through complex performance. As McGaw [4] 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 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. (p. 3)*

Performance-based assessment is not new. Oral and laboratory examinations have been used in schools and universities for over a century. In many industries performance-based assessment approaches are used (e.g. pilots). In many high-stakes courses in developed countries performance is, and has been, assessed using observation, interview, portfolio or recording (e.g. USA, UK, Denmark). For example, a recent review of assessment methods in medical education [5] outlines performance-based assessment of clinical, communications and professional skills using observations, recordings and computer-based simulations. In secondary schooling, there has been a history of performance-based assessment in some high-stakes courses, for example in the Arts, but this has been limited by the costs involved in collecting the evidence of performance and difficulties in ensuring reliable and valid results.

The general aim of this study was to explore the potential of various digitally-based forms for external assessment in an Engineering course in terms of manageability, cost, validity and reliability. The problem being addressed was the need to provide students with assessment opportunities in new courses that are on one hand authentic, where many outcomes do not lend themselves to being assessed using pen and paper over a three hour period, while on the other hand being able to be reliably and manageably assessed by external examiners. That is, the external assessment for a course needs to accurately and reliably assess the outcomes without a huge increase in the cost of assessment. The main research question was: *How are digitally based representations of student work output on authentic tasks most effectively used to support highly reliable summative assessments of student performances for courses with a substantial practical component?*

2. Method

The research design can be described as participative action research with participants contributing to development through evaluative cycles. As such this required an analysis of the perspectives of the key groups of participants (teachers, assessors, students) with data collected from each group. These data were compiled into case studies within a multi-case approach [1] in which each case is defined by one class. This approach allowed for refinement and further development of findings based on multiple instances of the same phenomenon under different conditions [7]. Therefore, this study largely employed an ethnographic action research evaluation methodology using interpretive techniques involving the collection of both qualitative and quantitative data.

A range of types of quantitative and qualitative data were collected including observation in class, a survey of students, a survey of the teacher, interviews with the teacher and a group of students, student work output from the assessment task, and the assessment records of the teacher. These data were analysed and used to address the research questions within a feasibility framework consisting of four dimensions:

Manageability (Can the digital assessment be managed in a normal classroom?),
Technical (Can existing technologies be adapted for digital assessment purposes?),
Functional (Is the digital assessment data reliable and valid?), and
Pedagogic (does a digital assessment support and enrich students' learning?).

The digital assessment tasks outcomes were marked independently by two external assessors, using detailed sets of criteria, which were represented as rubrics, and linked to the Engineering 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 collection of work for each course was marked using the method of comparative pairs, and these results were again compared against the results from the other forms of marking.

The Engineering course was selected as part of this research because it is a completely new course and its outcomes include processes and practical performance. The assessment task was implemented in five schools where Year 11students were studying Engineering Studies. In the first year a wide range of technologies were trialled in order to determine their appropriateness and success within the digital assessment context. These included a web-based digital portfolio system created in FileMaker and also a stand-alone system making use of net book PCs. In the second year the task was refined and the technology streamlined in order to determine the issues of scalability. 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.

3.1 Year1

The task was designed with the teachers who were currently teaching Engineering, and proceeded through a number of meetings and online refinement of the elements of the task. It involved a series of specified activities which took the students from a design brief to the construction and evaluation of a model over a period of 3 hours. Each activity was timed, so all students had the same specific time frame in which to complete each activity.

The task was presented to the students in the following manner:

> Families living in remote areas in developing countries have no access to town amenities such as power or water. They collect water in dams or tanks and use local material as fuel for heating and cooking.
> The purpose of this task is to design and model a solar water heater for a family living in a remote area of a developing country who collects their water in a tank adjacent to their house.

A portfolio template was developed using Filemaker Pro, which included instructions for students, and spaces for their input of either text, voice, sketches, pictures or videos.

Other Template pages required students to list the principles of appropriate technology, make a webcam presentation about the features of their design, evaluate other students' sketches and respond to peer evaluation of their own design.

The students were required to do some sketching of their ideas on paper, and then they took a picture of their sketch to include in their e-portfolio. A paper template was prepared for this purpose, folded to promote the sequence of activities required, and printed on 2 sides of an A3 sheet.

An initial visit was made to each school in order to discuss with the teacher the most appropriate physical facilities available for the implementation of the task. The task involved students working on computers, and also needing some desk space for the sketching and modelling aspects of the task. In all cases the task was implemented in an area where there was ample power outlets for the computers and peripherals, and also desk space for sketching and modelling with materials.

It was not necessary to visit the school for a trial of the hardware and software because the complete and self contained system was brought to each school by the researchers. This included all modelling materials, to ensure that each student had access to the same volume and type of materials so valid comparisons could be made between schools.

All teachers had to make timetable adjustments to enable the relevant groups of students to be available for the 3 hour block of time, which in realty was 3½ hours including the initial instructional and set up time. At most of the school sites, this period ran across lunch time and so the students were given a 20-30 minute break from the activity at an appropriate time.

A process was established, before implementing the task, to familiarize the students with the hardware and software. This involved students first of all setting up the computer and peripherals (mouse, web cam and memory stick) to ensure it was all functioning, and then going through each of the main elements of the task such as entering text, taking a picture, using both the external and integrated web cams and saving images into the portfolio template. This process was managed through a power point presentation by the facilitator, which included instructions and screen captures. When students had successfully mastered each element, the task was begun.

Students were allocated an ID number when they were issued with their USB thumb drive. They used this number to identify their portfolio, sketch sheet and model. In the collection of data then the student name does not appear. The students were then paced through the task activities, recording their output in their portfolio.

The three-hour exam was carried out without the use of any school resources, both the material for modelling and the ICT equipment was brought to the school by the researchers. Each student was allocated a mini computer (ASUS EeePC) for use in completing the engineering task. The battery on these computers lasts for about 3 hours, and because that is the length of the task, it was judged as inadequate for the time period hence the power cables were used, each computer was also accompanied by an external camera and mouse.

A FileMaker Pro database was used to develop the portfolio template and was loaded onto a USB memory stick. This was the mechanism used to capture all the digital student work: text, voice, pictures and video. At the end of the task, in the last page of the student portfolio, all the data entry boxes from the portfolio were collated on the one screen so students and researchers could make sure all pages had been recorded correctly. The memory sticks were removed from the computers upon completion of the task, and then later each of these individual student databases was combined to produce a master database of all student work which was uploaded to a web server.

The camera was the only tool used in the data collection process that could be improved. The integrated web cam was appropriate to record the student presentations but the USB camera posed some difficulties. This was used by the students to take pictures of both their sketches and their models. It worked well with the models because the camera could be easily moved to the appropriate angle to illustrate a specific feature of a model.

Because the camera focus could be adjusted, and the focal length was critical in ensuring sketches were recorded in adequate clarity, the fact that it was hand-held did not always result in crisp representations.

There were five schools, six teachers and five classes of Year 11 and one class of Year 9 students involved in the project who were doing the Engineering course. For each case the survey of students was done immediately on completion of the performance examination.

Broadly, it sought students' opinions on the examination itself, use of computers and other digital devices, attitudes to using computers and facility with computer applications. The questionnaire consisted of 45 closed response items and two open-response items.

The exam outputs for the 66 students were uploaded to the online repository. The students' work was marked by two external assessors using a standards based rubric. At the same time each teacher marked their students' work using his/her own method. The two external assessors marked the student work on the criteria developed for the assessment task using analytical 'marks' and converting these to scores using Rasch Modelling through the RUMM software package.

The analytic marking tool was developed from the Engineering Course of Study. The relevant outcomes were selected, and links were made with the engineering examination tasks. Each outcome was allocated a value, and then descriptors were developed to indicate a high, medium or low mark for each. The tasks which addressed each outcome were linked to the outcomes.

The marking methodology of comparative-pairs was also used. The marking tool designed for this purpose displayed two students work side-by-side, with the recording of the marker's choice as to which one was best. The tool was developed using FileMaker Pro and deployed on the Internet. This comparative pairs method of marking was used with five assessors each making 276 comparisons. One holistic and three specific assessment criteria were developed for the comparative pairs marking from the criteria previously developed for the task, and assessor was required to make the four choices for these criteria. The criteria were:

- **Holistic Criteria**: Progression of ideas and knowledge of materials clearly communicated within the design context. These holistic criteria related to the students ability to progress from their initial idea, in response to a range of stimulus and activities, to a satisfactory solution in a manner that clearly communicated the rationale for doing so.
- **Specific Criteria 1: Communication and progression of ideas.**
  This criterion dealt with the students' ability to clearly communicate their ideas through sketching, talking and writing.
- **Specific Criteria 2: Materials**.
  This criterion dealt with the students' selection of materials appropriate for their design.
- **Specific Criteria 3: Awareness of context**.
  This criterion dealt with the students' awareness of the design context – a remote area of a developing country – and the need to implement appropriate technology solutions in this context.

There was a strong and significant correlation (0.780 $p<0.01$) between the mark generated by comparative pairs marking and the mark determined by analytical marking. As might be expected, the 3 separate criteria (pairs marking) are also strongly correlated with the average analytical mark. There was also significant correlation between the teacher's examination mark and the pairs marking.

3.2 Year 2

The task evolved from that used in year one but still specified a series of activities which took the students from a design brief to the construction and evaluation of a model over a period of 3 hours. As in year one each activity was timed, so all students had the same specific time frame in which to complete each activity.

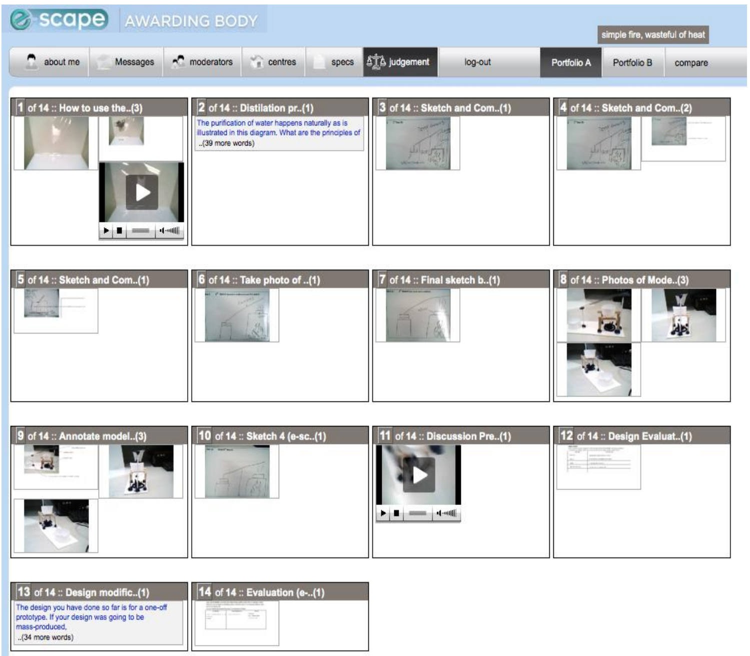
The task was presented to the students in the following manner:

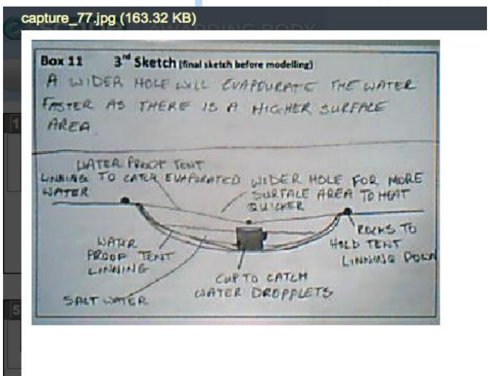
> The context for this task is a family camping at a remote beach. They have no transport, and have run out of fresh drinking water. There is no water around and so they need to invent a process to make seawater into drinking water. They have no power, so must depend on the heat and energy from the sun.
> You are to sketch and then model a system for turning seawater into drinking water.

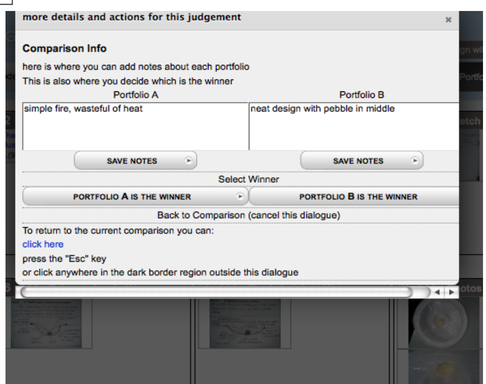
A portfolio template was developed using Tag Learning's e-scape web based assessment management system which included instructions for students, and spaces for their input of either text, voice, sketches, pictures or videos. The paper workspace template used in the first year was again used. As in the first year, other template pages required students to list the principles of appropriate technology, make a webcam presentation about the features of their design, evaluate other students sketches and respond to peer evaluation of their design. Contact with schools and other preparations were done as in year one. The e-scape portfolio environment was delivered either via the school's computers and the Internet or in some schools the Eee PCs used in year one communicating with a classroom-based server via wireless networking. This equipment was provided by the researchers and was used where the school's Internet or computers proved problematic.

The material for modelling and (in some cases) the ICT equipment was brought to the school by the researchers. Students were either allocated a mini computer (ASUS EeePC) for use in completing the engineering task or used a web browser and a classroom computer. A separate web camera and stand that allowed the taking of pictures was provided in both cases. The camera stand was developed as a result of the focus issues identified in year one.

The exam marking was carried out much as for year one, with multiple markers undertaking a comparative pairs style ranking procedure. The year two was undertaken within the e-scape portfolio environment which allowed markers to make comparisons between portfolios that contain various media types including video and audio and graphics. The figures below show the environment itself and the judgement screen. There were 104 students and after 13 marking rounds by 8 examiners the reliability coefficient was 0.898.







3.3 Year 3

The third year of the study involved eight Engineering teachers and the eight classes of Year 11 and 12 students they taught, with data collected from a total of 94 students. The students were studying the Engineering Studies course at either Stage 1 or Stage 2 levels.

The assessment task required a two-hour performance computer-based exam that was implemented for each class using the e-scape exam management system through the MAPS portfolio system (provided by TAG Learning Inc.). The examination was either run live from a website (online), from a wireless intranet within the classroom, or off USB flash drives on each student's computer.

The exam, facilitated by a researcher (or invigilator) and the teacher, was implemented with relatively few technical difficulties evident. This was largely due to the availability of the three modes of delivery. The exam was focussed on the design and development of a solution to a problem of producing drinkable water from seawater with a limited range of materials and using sunlight as the power source, similar to the second year of the project. Students were guided through a series of tasks for which stimulus material was presented. In online or wireless methods of examination delivery, the teacher controlled the student progress through the tasks, while in the flash drive method the student had control of their own progress with the teacher recommending the movement to the next activity. The same assessment task was implemented for each case (school) with the key difference to years 2 and 3 being that the 3D modelling aspect of student ideas was dropped on the advice of the Curriculum Council advisors.

The major variation between schools was the mode of the delivery (online, wireless or flash drive), but this variation did not alter the presentation format of the examination to the students. With the online and the wireless systems, the student work was automatically uploaded to the external server for the MAPS portfolio system. The students who worked off the USB drive had their work uploaded later to the server for marking. This required inserting the USB flash drive into a computer that has Internet access and logging into the MAPS server using the student login. As a result this method of delivery was only used when it was not possible to use either of the other two methods due to network firewall restrictions, Internet bandwidth constraints or software incompatibility.

As in previous year, a range of data was collected related to each teacher and class involved. These data included: observation of the class completing the exam; a survey of students; interviews with students, teachers and assessors; and scores generated by three methods of marking.

The three methods of marking used were: external analytical marking by two expert assessors; comparative pairs judging by some of the teachers and other expert assessors; and marking by the teacher for their own purposes and using their own methods. These data were analysed both for each case study (teacher and class) and for the combined samples of all teachers and students.

Overall Conclusions

As the study has progressed the technologies have been rationalised and streamlined so that they may be scaled up in larger trials.

While the technology has varied during the study, the examination has consisted of a design task that was scaffolded into a number of timed activities, progressing students from vague ideas to a realistic solution. Students were paced through each activity, recording their output in the form of a portfolio.

Although there were not always significant correlations between the markers scores using the analytical rubric based approach, overall the correlation was low but significant at 0.43. There was little correlation with any of the external assessors' marks and those provided by the teachers. The comparative pairs approach to marking provided a reliable set of scores (SI around 0.92) that was significantly correlated to the analytical marking scores ($r = 0.78$). There were similar outcomes for rankings on the marking approaches.

Overall the results of the third year of the study indicated that the benefits of the digital form of performance exam implemented outweighed the constraints. The assessment aligned well with the pedagogy of the Engineering Studies course, and most students enjoyed the practical design nature of the examination task.

The vast majority of students were enthusiastic about undertaking this form of examination. Apart from some disinterested students the vast majority were fully engaged with the task. Once the main technical issues were overcome it was relatively easy to invigilate. The students indicated a willingness to complete the exam on the computer and the ability to demonstrate their understanding through drawing, taking digital photos and commenting while taking video proved to be an influencing factor here.

This research demonstrated that a computer-based performance exam could be constructed for the Core content of the Engineering Studies and that this could be readily implemented in a large range of schools offering the course. However, in a number of ways minor improvements could be made to the structure, content and implementation of the exam.

References

[1] Burns, R. B. (1996). *Introduction to research methods.* South Melbourne, VIC: Addison Wesley Longman Australia Pty. Limited.

[2] Kimbell, R. (2010, April). *Project e-scape.* [Online] Retrieved from http://www.gold.ac.uk/teru/projectinfo/

[3] Lane, S. (2004). Validity of High-Stakes Assessment: Are Students Engaged in Complex Thinking? *Educational Measurement, Issues and Practice, 23* (3), 6-14.

[4] McGaw, B. (2006, May). *Assessment Fit for Purpose*. Paper presented at the Thirty Second International Association for Educational Assessment (IAEA) on Assessment in an Era of Rapid Change: Innovations and Best Practices, Singapore. Retrieved from http://www.iaea.info/papers.aspx?id=68

[5] Norcini, J. J. & McKinley, D. W. (2007). Assessment methods in medical education. *Teaching and Teacher Education, 23*(3), 239-250.

[6] Ridgway, J. McCusker, S. & Pead, D. (2006). *Report 10: Literature Review of E-assessment.* United Kingdom: Futurelab.

[7] Willig, C. (2001). *Introducing qualitative research in psychology adventures in theory and method.* Buckingham, UK: Open University Press.