1-1-2012

Do you want authenticity with that?: Online practical exams in an engineering course

Jeremy Pagram

P. John Williams

DO YOU WANT AUTHENTICITY WITH THAT? : ONLINE PRACTICAL EXAMS IN AN ENGINEERING COURSE

Dr JEREMY PAGRAM
CSaLT, Edith Cowan University, Perth, Australia
Associate Professor P JOHN WILLIAMS
CSTER, University of Waikato, Hamilton, New Zealand,

Abstract

This paper reports on part of 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. This paper focuses on the use of a computer managed examination in engineering studies. A computer-managed examination was designed that consisted of a design task that was broken down into a number of timed activities. Students were paced through each activity, recording their input in the form of a portfolio. Input consisted of text, graphics through a camera, video and voice. The exam outputs were uploaded to an online repository; the students' work was then marked by external assessors. The general aim was to provide students with assessment opportunities that were authentic, while being able to be reliably and manageably assessed by external examiners.

Introduction

The study, part of which is reported here specifically addressed a critical problem for the school systems in Western Australia (and beyond): the assessment of practical performance by traditional paper based examination methods and built upon the work of Kimbell (2004). Performance is a significant component of subjects such as foreign languages, art, physical education, engineering, drama and information technology. Traditionally, many of these subjects were summatively assessed through a three-hour written examination, which established a dichotomy between the essence of the subject and its assessment, and often promoted in appropriate pedagogies.

In this paper we focus on the use of a computer managed examination that was devised to support the assessment of engineering studies. We argue that the study advanced the knowledge base concerning the assessment of practical performance by developing techniques to represent practical performance in digital forms, to collate these in online repositories, and to judge their quality using a standards-based marking method and trialling a comparative pairs marking method. This paper focuses upon one component of the overall study, that of senior Engineering. Since the 1990s, low-cost, high-powered portable computers, and computer networks have increasingly appeared in schools. During that period many school systems have moved towards a more standards-based curriculum, and in keeping with this they have investigated methods of efficiently and effectively assessing. Many of the high-stakes senior secondary courses implemented in WA have a significant performance component, and are not able to be adequately assessed using paper-based methods. Therefore, it is important that other forms of assessment are considered, along with the potential for digital technologies to support this assessment.

The aim of this paper is to report on our investigation into the potential of various digitally based forms for external assessment in an engineering course under the criteria of manageability, cost, validity and reliability. Thus the problem being addressed was the need to provide students with assessment opportunities in new courses that are on the one hand authentic, while on the other hand, being able to be reliably and manageably assessed by external examiners. That is, code for external assessment to accurately and reliably assess the outcomes without a huge increase in the cost of assessment.

Overview

The research design can be described as responsive evaluation research with the feedback from participants contributing to development through evaluative cycles. As such this involved data collection from and analysis of the perspectives of each of the key groups of participants: teachers, remote assessors, students. These data were compiled into case studies within a multi-case approach (Burns, 1996) 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 (Willig, 2001). Therefore, this study largely employed a responsive evaluation methodology using the interpretation of both qualitative and quantitative data.

The quantitative and qualitative data collected comprise: Classroom observations, Questionnaire responses, Interview responses, and student achievement data (from the research assessment task) and from other school-based assessments. 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?).

Two external assessors, using detailed sets of criteria, which were represented as rubrics, and linked to the engineering course content and outcomes, marked the digital assessment tasks' independently. Correlations were made (for comparison purposes) between the assessments made by the two external assessors and also between the assessments made by 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 compared against the results from the other forms of marking.

The Engineering course was selected to be part of the overall study because it is a completely new course and its outcomes include processes and practical performance. The assessment task for engineering was implemented with Year 11 students in five schools. 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 (see Table 1).

Table 1. Summary of the scope of the study, timeline and related project activities

Phase (Year) and Scope	Project Activities
Proof of Concept (Year 1) Five classes Extended production exam Assessed by panel of assessors	Situation analysis including performance requirements, criteria and context. Design, creation, expert review, and testing of assessment tasks and instruments. Develop web-based repository. Digital representations collected by researchers with help from teachers and students. Online repository populated by researchers. Training and marking by assessors. Collect survey, interview and other assessment data.
Prototype (Year 2) Six classes Extended production exam Assessed by teachers and panel of assessors	Modification of assessment tasks and instruments, and online repository. Digital representations collected by teachers. Online repository populated by teachers/students. Direct capture for at least one course. Comparative pairs marking by assessors. Collect survey, interview and other assessment data
Scalable Product (Year 3) Eight classes Extended production exam Assessed by trained teachers and panel of assessors.	Evaluate and refine assessment tasks/instruments. Modify web-based repository for student work. Train teachers in assessment procedures. Digital representations collected by teachers. Online repository populated by teachers/students. Comparative pairs marking by trained teachers. Collect survey, interview and other assessment data.

Proof of Concept (Year 1)

The task was designed with teachers who were currently teaching Engineering, and proceeded through several meetings and online refinement of the elements of the task. It involved a series of specified activities that 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 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 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 two sides of an A3 sheet. 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. 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 modeling 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 power cables were used, an external camera and mouse also accompanied each computer.

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.

In retrospect, 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 t he students to take pictures of both their sketches and their models. However, 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.

In addition to the five schools of Year 11 students mentioned above, there was one class of Year 9 students, plus the six teachers, 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 Modeling through the RUMM software package.

Prototype (Year 2)

This task evolved from that used in year one but still specified a series of activities that 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 (Kimbell, et al, 2007), 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 EeePCs 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 modeling 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 figure 1 below shows the environment itself and the judgment screen.

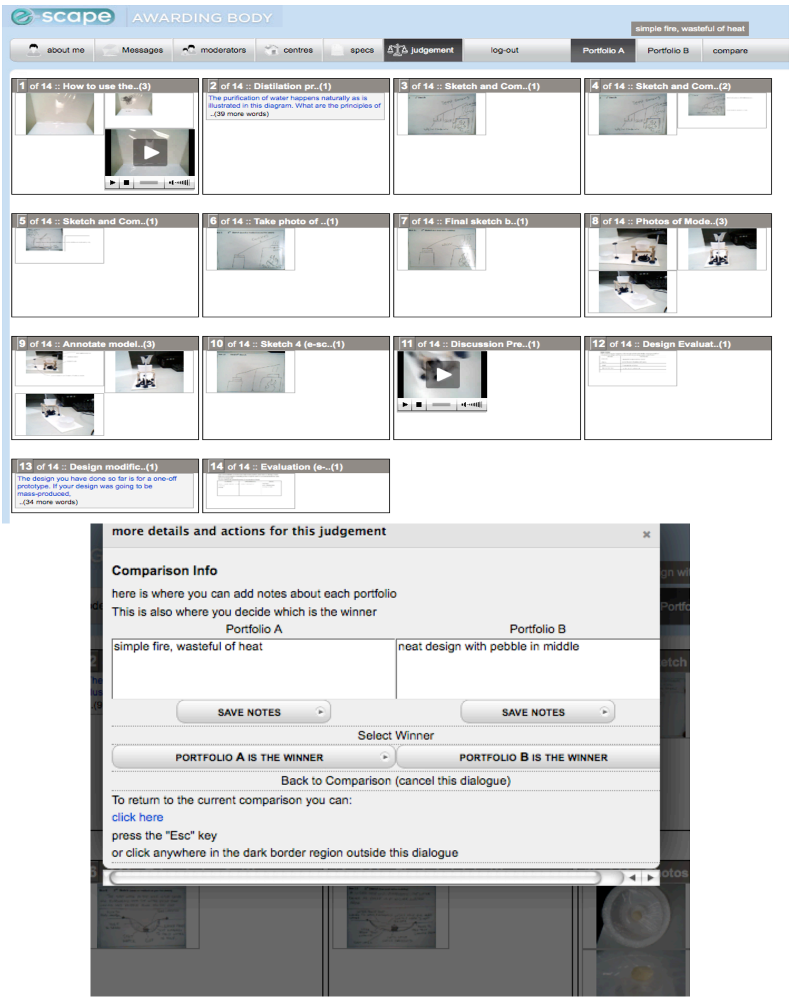

Figure 1. The exam environment (below) and the judgment screen (above)

Scalable Product (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 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 focused 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 modeling 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.

Discussion

By the third year the study had identified relatively few constraints to the use of the digital form of performance exam implemented in the sample of schools, and that the benefits outweighed the constraints. In all schools some adjustments had to be made to the school network system, either to allow the students to log on to the examination server or to install drivers for the cameras or the USB drives. In only one school was there difficulty organising access to a computer laboratory, and in this case the examination was conducted at the university. The only constraints in the marking processes concerned the need to upload the students' work from the USB drives, and this resulted in portfolio pages appearing out of sequence when presented to the assessors.

Some students indicated needing more time on specific tasks within the examination, and in those cases where the students were working off USB drives and pacing themselves through the examination, they all finished the examination within the allotted time. In a very few cases, students had trouble interpreting the instructions or the terminology used. A small number of students did not appear to take the exam seriously in those schools where the teacher did not include the examination in their school-based assessment.

The vast majority of students were enthusiastic about undertaking this form of examination. Apart from a couple of 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 despite many of them expressing concerns about using the approach. The ability to demonstrate their understanding through drawing, taking digital photos and commenting while taking video proved to be an influencing factor here. The ease to which the students could edit their responses and reflect upon their achievements during the process was also mentioned highly in their list of praise.

An important finding of the study was that in all cases the required IT infrastructure was adequate to ensure the assessment task could be completed to an acceptable level. There were some limitations that had to be overcome in USB port access and browser and flash software. While Internet access speed was a problem in four schools, alternative solutions were successful.

Detailed analytical marking criteria were developed for the assessors based on the task(s) and the standards framework for the course. These were represented as rubrics. The results of analytical assessor marking tended to be only moderately correlated. The scores and rankings from analytical marking were significantly but only moderately correlated with the scores and rankings from comparative pairs marking. There was generally moderate correlation between the results of marking by the teacher and the analytical marking and/or comparative pairs marking. There were greater differences between the analytical marking assessors and also between the two methods of marking than for the previous years. Differences between analytical assessors were not correlated with differences between the two methods of marking. A detailed analysis of a number of portfolios in this category did not reveal an explanation for this.

Conclusions

The study used a four-dimensioned feasibility framework to investigate the effectiveness of the assessment. Conclusions are summarised below under the headings of this framework.

Manageability Dimension

- There were few logistical difficulties except the booking of computer laboratories.
- There were some logistical difficulties concerning the management of student time to ensure they completed all requirements.
- Invigilators for computer-based examinations would need specific knowledge and skills to manage these forms of assessment.
- 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 are needed to familiarise students with the assessment tasks.

Technical Dimension

- 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 would be necessary.
- Audio or video recorded responses to assessment task items would need to be stored locally (in addition to any attempted streaming of responses to a server) to minimise data loss.
- Any online examination system would need to be adaptable to a variety of technical environments across the variety 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).
- Some degree of maintenance is required to prepare student work for assessors.

Functional Dimension

- 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
- 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 both students and teachers appreciated this.
- Students would need to be prepared with practice tasks of the kind to be implemented in an online assessment task to ensure students could perform optimally during the task.

Pedagogic Dimension

- Typically, students liked doing the practical work involved in the digital assessment tasks and, 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.
- 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.

Acknowledgments

This Research was conducted by a team from CSaLT at Edith Cowan University, under the leadership of Associate Professor Paul Newhouse.

Investigators: Associate Professor Paul Newhouse, Associate Professor John Williams, Dr. Jeremy Pagram, Dr. Alistair Campbell, Dr. Martin Cooper

References

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

Kimbell, R. (2004). Design & Technology. In J. White (Ed.), *Rethinking the School Curriculum: Values, Aims and Purposes.* (pp. 40-59). New York & London: Routledge Falmer.

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.

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