TENZ 2009
Napier, New Zealand, October 6-8, 2009

ASSESSMENT OF STUDENT PERFORMANCE IN ENGINEERING

Assoc Prof P John Williams
Edith Cowan University
University of Waikato

Introduction

This report is the result of the first year 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 and supported by an Australian Research Council grant. The study commenced in January 2008, will be completed by 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. This paper will focus on Engineering Studies.

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' (Lane, 2004; Ridgway, McCusker, & Pead, 2006). 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 (2006) 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. (McGaw, 2006, p. 3)*

Performance-based assessment is not new. 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. 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 (Norcini & McKinley, 2007) 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 highstakes 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.

Problem and Research Question

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. 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?*

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 (Burns, 1996) in which each case is defined by one digital form of assessment in one class for one course. 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 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?).

These digital assessment tasks were marked independently by two external assessors, using detailed sets of criteria, which were represented as rubrics, and linked to the 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 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 Case

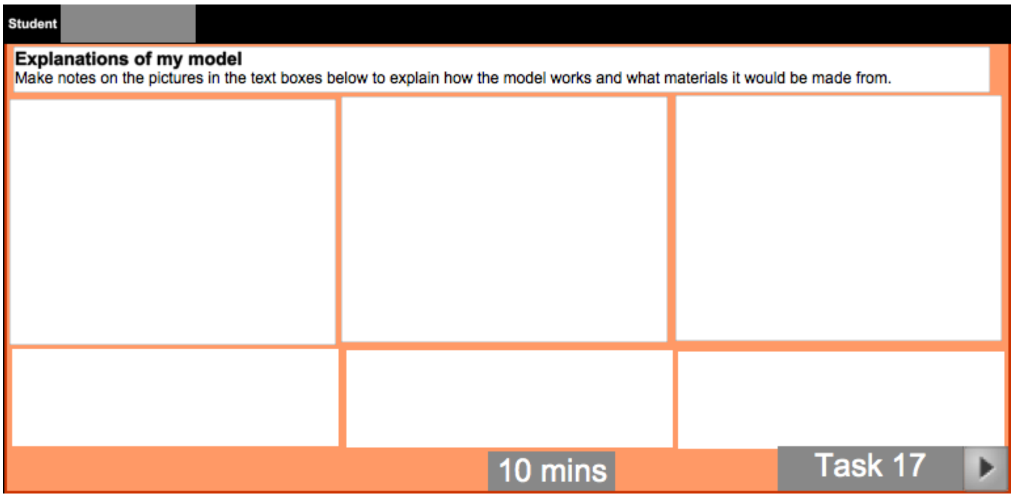
The Engineering course was selected as part of this research because it is a completely new course and its outcomes include processes and practical performance. The assessment task was implemented at five schools with a total of five classes (two at one school) of Year 11 and one class of Year 9 students studying Engineering Studies.

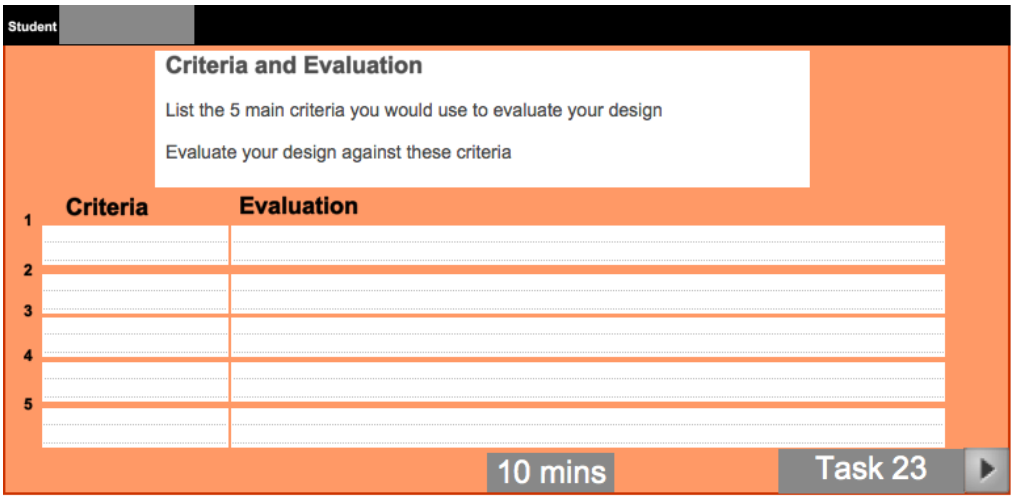
The Assessment Task

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.

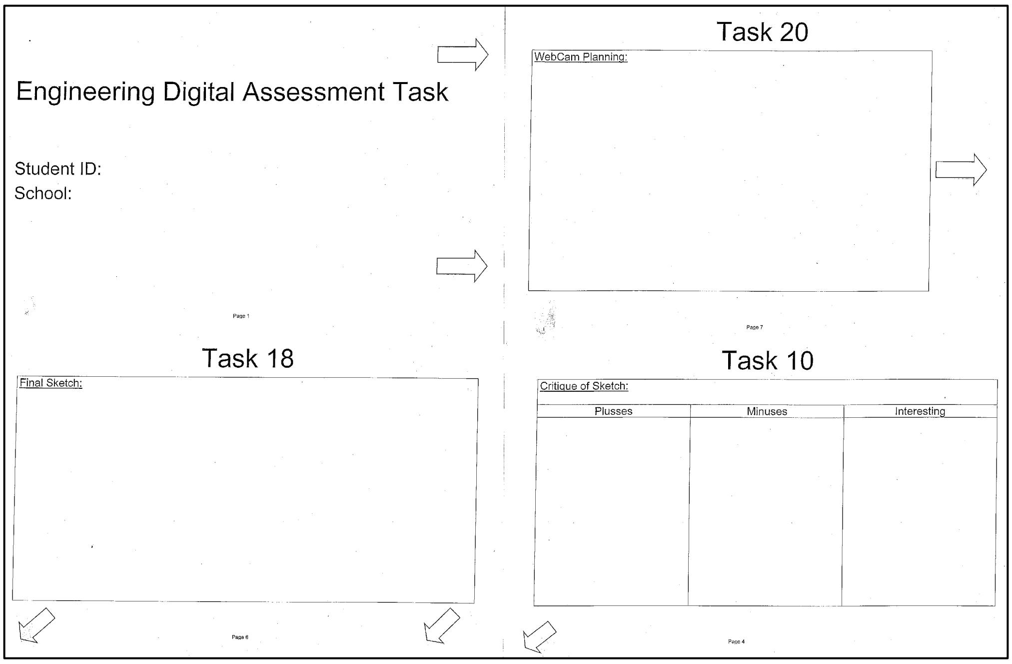
A portfolio template was developed using Filemaker Pro which included instructions for students, and spaces for their input of either text, voice, sketches, pictures or videos. The following are two sample template pages.





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.

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. For example:



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. No students had any difficulty with these tasks, in fact they were invariably running ahead of the facilitators instructions and exploring all the elements of the hardware and software. 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.

Technologies
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 additionally accompanied by an external camera and mouse.

The ASUS eeePC with peripherals attached

A FileMaker Pro database was used to develop the portfolio 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.

There were no serious technical impediments to the conduct of the task in any of the schools. This was anticipated in this context where the researchers had control over all aspects of the technology. Even the power-point projection that was used to manage the task was done on a laptop and data projector that was brought into the classroom by the research team. Both the hardware and the software were quite robust and when transported from school to school worked well.

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.

A student using the USB camera to take pictures of his 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.

Marking

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 the teacher marked the 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.** This 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 criteria dealt with the students ability to clearly communicate their ideas through sketching, talking and writing.
Specific Criteria 2:	**Materials.** This criteria dealt with the students selection of materials appropriate for their design.
Specific Criteria 3:	**Awareness of context.** This criteria 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.

The RUMM software provided a summary of results that ranked from best (most number of times preferred) to worst (least number of times preferred). A Separation Index was calculated as an indicator as to whether or not the exemplars were sufficiently diverse in quality to assure a broad enough range for the purposes of comparison. It is given as a number from 0 to 1. Values closer to 1.00 are more desirable. If the value is close to 0.00 (up to about 0.3 or 0.4) the range is too narrow. If it is above about 0.7 the separation is reasonable and if it is above 0.8, the separation is good. The Separation Index for the Holistic criteria was 0.917 indicating a very good spread of quality in the exemplars, and the group reliability was 0.992.

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.

Conclusions About Marking

The amount of time taken by the expert assessors in the analytic marking averaged about 30 minutes per student. For the comparative pairs marking assessors took an average of about 2 minutes per comparison.

Comments of Assessors

The assessors were asked for feedback on the process of marking in terms of the following questions:

(1) Assessment task - appropriateness and quality

- suitability of the task for the course: *The task did test modelling and design aspect of the course in a practical way but did not test the practical skills. It tested mainly the design nature of the course, not a practical engineering test.*
- engagement with students: *Yes my students did engage in the task, it was used as an assessment point.*
- breadth to allow all students to demonstrate performance: *it was very difficult to demonstrate performance due to time limitation*
- limitations of the task: *Only tested design skills, materials used not real*

(2) Marking process - rubrics and online tool

- ease of accessing student work: *Very time consuming , moving between screens (4) 30 to 40 mins start down to 20 mins, access easy, monitor size an issue*
- ease of entering judgements/marks: *entering judgments/marks easy*
- suggestions for improvements: *Larger monitor, improve moving between screens, improve quality of student drawings, Due to videos needed work on a PC.*

(3) Quality of student work

- general standard of work: *Nothing outstanding, some good work*
- opportunity for students to demonstrate quality: *Yes, it was about quality of thought on design, not workmanship*

Conclusions About Students

There were 48 students who completed the survey. The minimum, maximum, mean and standard deviation were calculated for each closed response item using SPSS. Responses to the open-response items were tabulated to assist in drawing out themes.

Most students had done a design project on a computer before and so were used to the process. They found using the computer easy, and that it was good quick way to record their design and modelling ideas, and an appropriate tool to use in the compilation of their portfolio. All the students used a computer, MP3 player and mobile phone at home, and most used a digital camera and game console. At school they used a computer for an average of about 50 minutes each day, though with significant variability within the group, from no time to over 5 hours. Most students were confident with computers and felt they were good at using them. In reference to a range of computer software and equipment, the students felt least comfortable with databases, video editing and web page authoring, and most comfortable with PowerPoint, email, the internet and digital photography. Overall they indicated a high self-assessment of their computer skills.

Conclusions from Teachers

In general the teachers were very supportive of the approach used in the assessment task and felt that it complemented the nature of the subject and that their students enjoyed it, although some students who were not used to designing struggled with that aspect of the task. However, they all suggested improvements, most commonly to the capture of the student sketches which made the marking difficult, and maybe not truly reflective of the students capability. They agreed that it was important for students to have a break during the 3 hour activity. One teacher felt that the students should be provided with the assessment criteria prior to participating in the assessment task.

Findings

A summary of findings from the Engineering case studies was compiled based on the Feasibility Framework and including a summary of the constraints and benefits of the form of assessment used.

Manageability

- Set up and take down easy.
- Ideal rooms had plenty of power outlets and bench space for computers and for 3-D modelling.

Technical

- No technical difficulties.
- No dependence on school infrastructure.
- Hardware generally reliable – 1/60 computers did not function appropriately.

Functional

- In all cases most students readily perceived the assessment task to be authentic and meaningful for their course, and engaging.
- The task was structured to permit a good range of levels of achievement to be demonstrated, but one teacher felt it could have been broader.
- While correlations between cases were not consistent, overall correlations were significant, and strong between analytic and pairs markers.

Pedagogic

- Typically students liked doing an assessment task in this form.
- The assessment task matched the typical pedagogy for the course, with the exception of one school where the students had not done much design.
- All students preferred the assessment task than a paper-based exam.

Constraints

- No significant constraints.
- Some found the size of the mini-computer a little constraining when typing.
- The resolution of the web-cam constrained the clarity of the images.

Benefits

- Typically students engaged positively with the task.
- The nature of the task is more aligned with the methodology of the subject than the traditional paper test.

Summary

The study involved five engineering teachers and their year 11 students, and one teacher's year 9 students in five different schools. In one school, two teachers and two classes participated. In total from the five schools, 66 students participated in the study.
The examination 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. The instructions to the students and input by the students were by way of a USB thumb drive in a mini computer. Input consisted of text, graphics through a camera, voice and video through a webcam.

Almost all students preferred the assessment of their engineering performance through this means rather than a paper examination. They found the experience engaging and enjoyable, and felt that it more accurately reflected the nature of the engineering course they were studying. Many students seemed to have a natural affinity with the range of technology used in the examination.

Their two main concerns were with the size of the keyboard on the mini computer resulting in difficulty in typing, and the low resolution of the webcam resulting in poor representation of their sketches. The difficulty in interpreting the development taking place in the series of sketches because of this poor resolution was also highlighted by a number of the assessors. The assessors commented that it seemed that those students for whom this examination was a part of their school based assessment were more focussed and treated the exercise more seriously and consequently achieved better results.

The spread of marks would seem to indicate that the task provided students with adequate scope to demonstrate their competency, but one teacher/assessor felt that a different task would increase this scope.

The timing of the activities within the three hour task generally seemed to be appropriate, although some of the faster students were frustrated at having to wait before moving onto the next task, and used this time to circumvent the structure of the portfolio and enable them to jump ahead. The modelling of their design idea was allocated 40 minutes, and many did not complete their model to their satisfaction within this time.

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.

As a result of the success in 2008 it is recommended that for 2009 a very similar assessment task be used with the only major changes being improvement of the use or resolution of the camera, and maybe more individual control by the students of the amount of time they spend on each aspect of the examination task. The assessment rubric in 2009 will cover the course content rather than the course outcomes.

In 2009 a move will be made in some schools to use the technology within the school as the platform for the examination rather than bringing the technology into the school by the

research team. This is in line with the 2010 goal of the project to increase the sample of schools involved to be more representative of all schools. Once again teachers will be strongly encouraged to include the examination task within their school programmes as part of school-based assessment.

REFERENCES

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

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

McGaw, B. (2006). *Assessment to fit for purpose.* Paper presented at the 32nd Annual Conference of the International Association for Educational Assessment., Singapore.

Norcini, J. J., & McKinley, D. W. (2007). Assessment methods in medical education. *Teaching and Teacher Education, 23*(1), 239-250.

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

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