Title: **Digital assessment: assessing what counts, the performance.**

Authors' Names: Hendrati Nastiti, Edith Cowan University , Australia
h.nastit@ecu.edu.au
Dr Alistair Campbell, Edith Cowan University, Australia
a.campbell@ecu.edu.au

Abstract
Future generations will look back at current assessment practices and wonder why it took so long to replace pen and paper assessment with digital assessment. Digitization of the assessment process, from student work to the recording of marks, is being done in education now but haphazardly and often only as a replication of the paper assessment. The desired goals and aims of assessment have rarely been fully achieved with pen and paper assessment. Could digital assessment finally achieve them?

This presentation will firstly provide an overview of current research into Digital Forms of Assessment (DFA) from the capturing of student authentic performance to high stakes comparative pairs marking. Followed by the research results and findings of the digitization of a 3-hour Year 12 exam in Engineering Studies that captured the student's performance overtime; in text, drawing, pictures and video. The teachers and students were surveyed pre- and post-exam to obtain their use and attitudes to ICT and to the digital exam. The student digital performance results were assessed in three ways, analytically by the teacher, by two external markers that used a digital marking rubric and by comparative pair marking. The reliability, validity, manageability and scalability are discussed.

Key words:
Performance assessment, digital assessment, ICT, authentic, innovative

Introduction

Future generations of students and teachers will look back at current assessment practices and wonder why it took so long to replace pen and paper assessment with digital assessment. The desired aims of learning have rarely been achieved with pen and paper assessment, because this type of assessment cannot capture the rich, complex learning experience. In addition, the traditional marking methods employed in the assessment process have been highly labor-intensive, involving many processes that are semi-skilled, time consuming and costly. The digitization of the assessment process offers the possibility to drive assessment and, thus, teaching and learning, in positive directions. Digitization allows the capture of authentic student work and performance within the school environment, efficiently, reliably and at minimal cost.

Digitization of the assessment process is already occurring throughout the world in the field of education; however, this process is often taking place by default and haphazardly, and the limited aspects that are being digitized are often only a replication of the paper process. This transition from an analogue to a digital assessment environment has resulted in many ludicrous and un-educative situations, such as Year 12 students having to stop using digital tools and start practicing handwriting, as their final high-stakes exams are paper-based exams. Similarly, teachers word-process the marking keys but then print them, handwrite names and comments, and record the marks manually.

The area of digital assessment research is not as well developed as other areas of digital education and, in this transitional period, research is needed to investigate how digitization can enhance the goals of education. In this presentation, I will give an overview of the current research being undertaken at the Centre for Schooling and Learning Technologies (CSaLT) at Edith Cowan University into Digital Forms of Assessment (DFA), and I will be discussing the findings of the digitization of a Year 12 exam in Engineering Studies. The discussion will cover the capture of the student's performance overtime using text, drawing, pictures and video, and the teachers and students' use and attitudes towards Information and Communication Technologies (ICT) and to the digital exam. The assessment of the student digital performance results, and issues of reliability, validity, manageability and scalability will also be discussed.

Background

Performance-based assessment is one of the oldest forms of assessment, and is essential in many types of jobs, as we would be reluctant to fly with a pilot or be operated on by a surgeon that has only learnt the theory of the job. Performance has been assessed in many ways, for example, through an observation or simulation, an interview or a portfolio of some kind. This type of assessment involves processes that are usually costly, especially at a school level. For example, in Western Australia, in Year 12 high-stakes courses such as Music and Arts, performance-based assessment takes place but to a limited extent due to the costs involved in managing the evidence of performance and the difficulties involved in assessing the work to ensure reliability and validity.

CSaLT, in collaboration with the Curriculum Council of Western Australia and supported by an Australian Research Council grant, is currently conducting a three-year study investigating the potential of digital technologies to represent student output from high-stakes performance-based tasks in four senior secondary courses: Applied Information Technology, Engineering Studies, Italian and Physical Education Studies.

In this study, the performance-based assessment was an 'extended production exam', the completion, under exam conditions, of one practical assessment task that incorporated a full set of processes (e.g. design process, scientific investigation) and centered on one major

scenario. Examples were found locally, nationally and internationally of performance on practical tasks being assessed through an extended production or a small project, under exam conditions; however, most did not involve the use of digital technologies. The most comprehensive digital example was that of Kimbell *et al* (Kimbell, Wheeler, Miller, & Pollitt, 2007) in the UK, where students spent two three-hour sessions on consecutive days working on a structured design activity for the production of a pill dispenser. All student work output was recorded digitally using a networked Personal Digital Assistant (PDA) device and stored on a local server.

Aim and Objectives

The aim of the three-year research study was to investigate the feasibility of using different digital media such as text, photos, graphics, audio and video to capture student performance and work within senior secondary courses under exam conditions, and to evaluate this process in terms of manageability, cost, validity and reliability. The digital exams were designed to provide students with authentic opportunities that could not be represented through a pen and paper exam.

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?

This study addressed this question by considering a number of subsidiary questions:

1. What are the benefits and constraints of each digitally based form to support the summative assessment of student practical performance in senior secondary courses in typical settings?
2. What is the feasibility of each digital form of assessment in terms of the four dimensions: technical, pedagogic, manageability, and functional?
3. Does the paired comparison judgments method deliver reliable results when applied to student practical performance across different courses?

Method

A participative action research design was used, with participants contributing to development through evaluative cycles. The key groups of participants were teachers, assessors and students, and data were collected from each group and compiled into case studies within a multi-case approach (Burns, 1996). Each case study was defined as one digital form of assessment in one class for one course; all case studies were then grouped and summarized by course of study. This approach allowed for refinement and further development of findings based on multiple instances of the same phenomenon under different conditions (Willig, 2001). Thus, this study mainly employed an ethnographic action research evaluation methodology using interpretive techniques involving the collection of both qualitative and quantitative data.

Qualitative and quantitative data were collected through a range of methods, including: observation within the class, surveys of students and teachers, individual interviews with teachers, group interviews with students, student work output from the assessment task, and teachers' marking records. Data analysis addressed the research questions within a feasibility framework consisting of four dimensions:

- Manageability – can digital assessment be carried out in a typical classroom environment?

- Technical – can currently available digital technologies support digital assessment?
- Functional – can digital assessment be reliable and valid, and comparable to other forms of assessment?
- Pedagogic – can digital assessment support and enrich students' learning experience?

All student work was marked in three ways: by their teacher; by two external assessors using an online analytic standards-referenced method of marking based on a set of criteria from the appropriate course content; and finally, by external assessors using an online comparative-pairs method. Correlations were determined for comparison purposes between the different methods of marking.

Engineering Studies

The Engineering Studies assessment task was implemented at five schools with a total of seven classes of Year 11 students in the Engineering Studies two-year course. The task was designed in collaboration with teachers currently teaching Engineering Studies, and consisted of a series of specific activities that took the students from a design brief to the construction and evaluation of a model over a 2-hour period. Each activity was timed, so that all students had the same time frame in which to complete the activity, and the research facilitator (invigilator) controlled the movement between activities within the examination classroom. Figure 1 shows the way in which the task scenario was presented to the students:

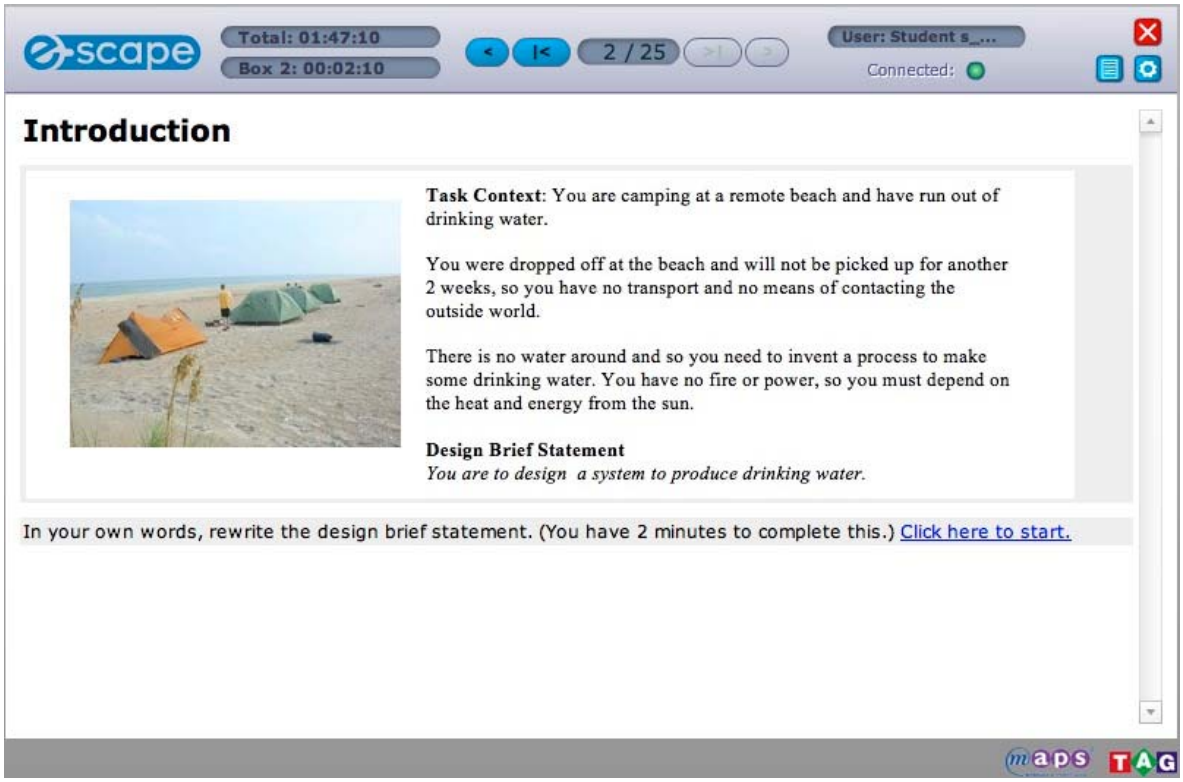

Figure 1: Task scenario presented to the students

'E-scape' was used to design and present the task to the students. This Internet-based software enabled the design of activities in which students can input digital data – text, graphics, voice and video. Stimulation material and instructions were included with each activity (or 'box') and were timed. The students were required to: 1) sketch their ideas on paper; 2) take pictures of the sketch; and 3) digitally annotate the sketch. This process went through four cycles as additional stimulus information was supplied. To complete the task, students were required to record a video-clip describing their sketch.

Technologies Employed

Each student worked on school computers or a supplied Netbook that had a webcam attached. Students logged directly into the 'E-scape' exam through a web browser, as shown in the images in Figure 2. The webcam allowed the student sketches to be photographed, and a short video to be recorded.

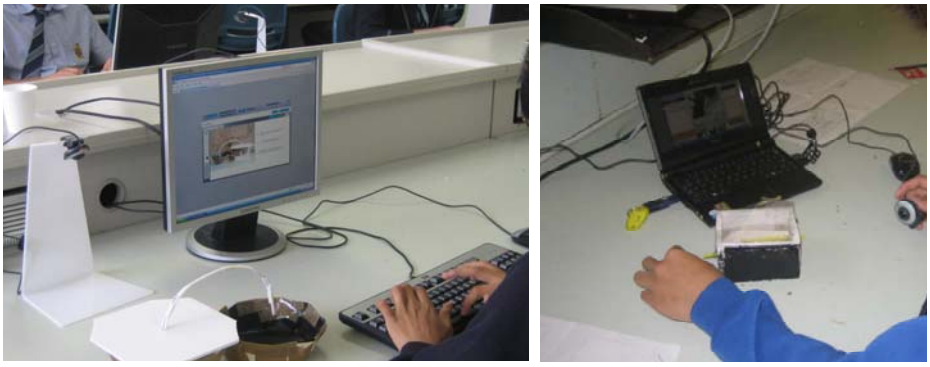

Figure 1: Technologies employed

Task Assessment

The engineering task was assessed in three ways: by the class teacher, who employed their own method of assessment; by two external assessors, who used an Internet-based analytical rubric based on the course of study; and, finally, by five external assessors, using a comparative-pairs marking method. The digital analytical rubric was designed at the research centre using FileMaker Pro, and the comparative-pairs marking tool (*Pairs Engine*) was developed by TAG Learning to complement its E-scape exam management system and resulting digital portfolios.

The marking of the engineering task using the analytical marking tool presented some problems for the assessors, as it was not possible to display the student work within the marking tool as the work was embedded within the online portfolio system. This meant that the link between the student marking window and the student work window had to be done manually.

The comparative-pairs marking tool combined all the main processes involved in marking using the comparative-pairs method (see Figure 3). The online system generated pairs of student work for each assessor to judge, and provided a facility for recording the assessor's judgement and related notes. The tool calculated scores and associated reliability coefficients using Rasch modelling while marking occurred, which meant that when an acceptable level of reliability was reached marking could stop.

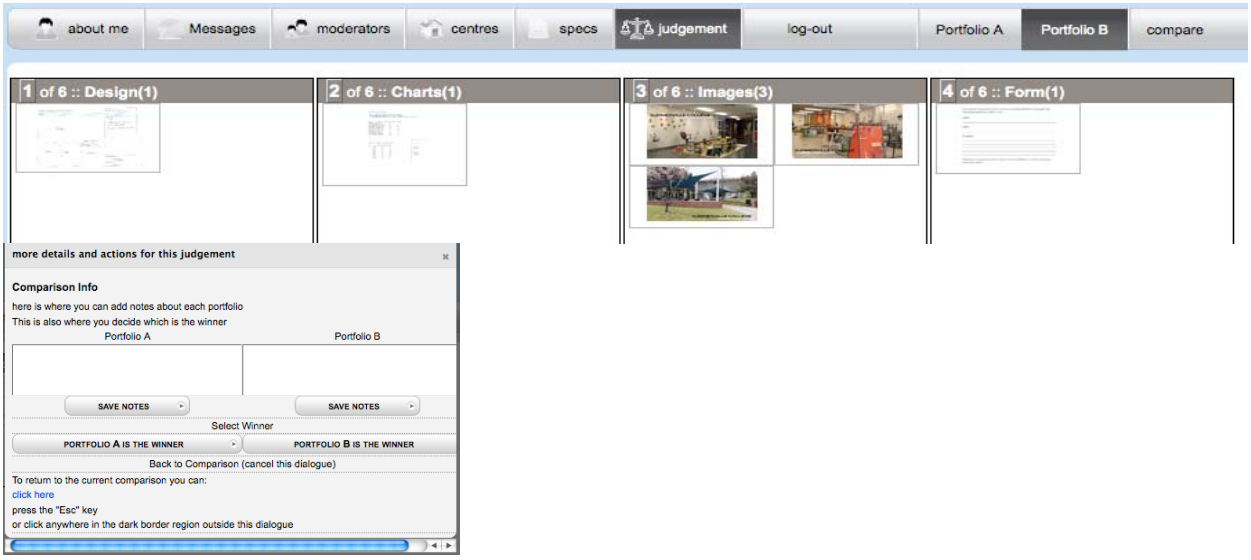

Figure 3: Screen displays from the Pairs Engine used for comparative pairs judging.

The comparison analysis of the three sets of marks showed that, overall, the correlation coefficients were very low and often not significant. This lack of correlation suggests a lack of reliability in the scores generated by the marking methods, which warrants further investigation to ascertain what caused this low correlation.

Student Response

The following is a summary of the student responses to the post-examination survey, which included questions related to their access to ICT and ICT skill level.

Most students reported having done a design project using a computer before, and thus were familiar with the process involved in the exam. Most reported using a computer, MP3 player and mobile phone at home, and had access to a digital camera and game console. They felt confident with using computers, and comfortable using databases, video editing and web page authoring, PowerPoint, email, the Internet and digital photography. Overall, students' self-assessed scores indicated a relatively high level of computer skills.

Students found that using the computer in the exam was easy, and that it was an effective and appropriate tool to record their design and modelling ideas. In the survey, students were asked to list the two best and worst aspects of the activity. Some of the best aspects highlighted by the students included the following:

- The digital exam allowed students to be more creative and present their best quality work;
- It was easier and faster to write comments;
- Students found a better connection between theory and practice; and
- Students generally enjoyed doing the task, and found it 'fun'.

Some of the least favourable aspects highlighted by students included the following:

- The time limit was considered too short for some activities;
- Access to the Internet was at times slow, and made it difficult to record work; and
- The low resolution of the webcam resulted in images being difficult to view.

Teacher Response

The teachers were generally very supportive of the digitized assessment task. They felt that it complemented the course objectives and that the students enjoyed the process. The general consensus was that they would like to see this type of assessment continue in their courses of study. The following quotes are representative of the teachers involved in the study:

- 'Designing a product is something kids struggled with and it is good to give them the extra time to do the job.'
- 'It has a huge amount of potential for practical subjects where a written exam is only half of the picture.'
- 'Capturing actual performances as apposed to text-based exam has more potential to being a farer representation of students' ability.'
- 'Students took this type of assessment very well and performance was a truer reflection of the student's ability.'

Summary of Findings

The summary of findings is based around the Feasibility Framework, and includes a summary of the constraints and benefits of this type of digital assessment.

Manageability issues were few within the schools: the school computers were used and, prior to the examination day, minor technical issues were resolved through discussions with the IT support staff, and the testing of the software.

Technical issues were few and the school infrastructure was reliable. The hardware was generally reliable; however, Internet access speed was an issue, which meant that occasionally the web page became frozen and students had to re-login.

The digital task examination process functioned well, with most students commenting that they felt that the assessment task was authentic and meaningful for their course, and was also engaging. The majority of the teachers felt that the task was structured to permit a good range of levels of achievement to be demonstrated.

Pedagogically, the findings have shown that this type of digital assessment can support and enrich the learning experience of students, and, on completion of the task, all students commented that they preferred it to a paper-based exam.

There were no significant constraints, except for the low resolution of the webcam resulting in the drawings being difficult to read, and slow Internet speed at some schools. The benefits were found to be that the students engaged positively with the task and that they preferred typing than writing because of efficiency, speed, neatness and clarity. The nature of the task and the ability to capture authentic student work was more aligned with the methodology of the course than the traditional paper-based exam.

Conclusion

The Engineering Studies digital examination findings mirror those of the other three courses under investigation. They demonstrate that with the technologies and infrastructure currently found within schools and the community, it is possible to capture authentic student performances in a wide range of situations in a cost-effective manner. The students in the study represented a typical range of abilities, and were able to demonstrate their knowledge and performance in a digital exam. The data, analysed within a Feasibility Framework, has

shown that there are many advantages to digital assessment over the traditional paper-based assessment methods.

References

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

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: Open University Press.