Imminence of e-assessment

How soon can we adopt online examinations for an HSC subject?

(Adapted from work submitted to Deakin University in 2010.)

Computer-based assessment comes in many forms and is adaptable.

Online instruments in higher education have not yet profoundly changed learning or assessment. It is reasonably easy and common to assess the ability to recall information and relate it to questions in a quiz structure. Learning Management Systems (LMS) are now able to include video or multimedia stimulus, adaptive or randomised questions, and provide automated feedback to candidates in real-time (Byrnes & Ellis, 2006 cited in Fluck, Pullen, & Harper, 2009). Multiple choice questions, cloze questions, and short answer questions that can be judged on keywords, can be marked automatically (BC Consulting, 2006, cited in Fluck et al.). Appropriate e-Assessment is preferred by most students, once familiar (Gilbert, Gale, Warburton, & Wills, 2009, p. 25).

A notable example is the e-AsTTle system in New Zealand, which allows teachers to select test items that have been validated for specific curriculum objectives, to make up an online test. Closed questions are marked automatically, but a teacher can include open questions, to be marked by the teacher (Hattie, 2010). This assessment is adaptive at two levels: the computer system varies the selection of questions in response to student performance, and teachers customise the selection of test items available. However, technological constraints can lead to over-reliance on Objective Response items and narrowing of the assessment (Gilbert, Gale, Warburton, & Wills, 2009, p. 26). Unfortunately, there is little capacity for computer administered tests to measure higher order thinking, creative work, or engagement in authentic tasks (BC Consulting, 2006). Isolated projects have attempted to assess participation in online environments such as Second Life or to utilise peer-ratings, which do seem to have the potential to assess creativity and critical thinking (Ripley, Tafler, Ridgway, Harding, & Redif, 2009).

Fundamental principles of assessment apply to e-assessment as much as to traditional forms, but implementation issues necessitate modified development processes. At this stage, the capacity for adaptive testing is largely under-utilised, and the impact of adaptive practices are poorly understood. The approach relies on predetermining ranked difficulty of all questions and all permitted sequences of questions, so that the computer can deliver a more or a less difficult question in response to student performance. The computer's control is usually achieved at the cost of preventing students from reviewing and changing their early answers. These constraints have led to a waning of interest in Computer Adaptive Testing (Gardiner, Holmes, & Leitch, 2009, p. 13). Penalty (negative) marking and time-dependent scoring are possible, but can have unexpected effects on learning and on validity of assessment. Processes used in developing electronic tests are not as well developed as processes for traditional tests (Gilbert et al., p. 31). A team approach may be necessary, to bring together a psychometric rigor and the content complexity of subject experts. Test development can benefit from team-authoring, trialling by teachers, interrogating a pilot cohort, mixing new questions into existing tests, and soliciting student comments (Gilbert et al., p. 33).

Gilbert et al. urge the use of psychometric principles and techniques, to evaluate the reliability of test items and testing methods. Possible interpretations of uniformly poor performance on a test item include technology flaws, poor question design, or lack of instructional validity (Gilbert et al., pp. 28-29). However, Item Response Theory (IRT) is only practically applicable in reasonably large test banks used with large candidatures (Hogan, 2007, p. 61). While software to accomplish the statistical processing are available (Partchev, 2004), most teachers are unlikely to determine the most appropriate form of IRT for their situation due to the subtlety of differences in IRT models (Linacre, 2003). Simpler steps can be used to exclude test items that are found (statistically) to be grossly less discriminating than the overall test.

Online exams have been observed to have some typical impact on students and teachers.

One trial piloted a computer-administered examination with third year undergraduate students. (Fluck, Pullen, & Harper, 2009). Notably, "the preference for examination medium appeared to be strongly related to successful prior student experience of using computers in assessment, with... an effect size of 0.621" (p. 518). "The student cohort was ambivalent about the introduction of computer based examinations." The researchers made recommendations regarding equipment (headphones and keyboards), equalising familiarity effects (typing speed and writing tools), and collection of completed examination scripts (avoiding networking if possible).

Test-mode effects in both directions have been found: various studies find that paper-based tests or computer-based tests of the same content systematically yield higher scores. An experimental attempt to dissect contributing factors found that scores on paper tests significantly exceeded scores on computer-based tests, for "egocentric" students – those who felt less collaborative (more directly competitive) and less engaged. Computer mode disproportionately strengthened the performance of high attaining students, presumably because they accommodated the new requirements more rapidly. (Clariana & Wallace, 2002). Equity implications must be considered, as students in urban fringes or with educated parents scored better on computer, in one large UK study (Ripley, 2007, p. 6), and socioeconomic correlations can be found in Australian Year 6 and Year 10 ICT test outcomes (MCEEDYA, 2008).

A test mode that is similar to lessons can be expected to result in higher scores, according to Transfer Appropriate Processing (TAP) theory (Bransford & Franks, 1976, cited in Clariana & Wallace, 2002). This raises the question of which test mode is most valid for a particular candidature, whose instruction is intended to meet a particular curriculum purpose.

Screen resolution, font design, or other physical factors may make screen-reading more tiring than reading print (Mourant, Lakshamanan, & Chantadisai, 1981; Wilson, 2001, cited in Clariana & Wallace, 2002). Randomising question order is easier and more common in computer-based tests, and may increase the cognitive load for students (Beaton & Zwick, 1990; Cizek, 1991, cited in Clariana & Wallace, 2002). Questions larger than one screen tend to result in lower scores on computer than on paper (Haas & Hayes, 1986, cited in Clariana & Wallace, 2002).

Teacher confidence in the stability of technology for computer-based testing is generally low. JISC "view delivery issues in e-assessment quality as hygiene factors, in that they may cause unsatisfactory assessments but cannot ensure quality in and of themselves," but they remain a preoccupying concern for many practitioners. This includes testing with the equipment and network conditions that students will experience, and minimising concurrent server load (Gilbert et al., pp. 27, 29, 34). Nonetheless, recent improvements in usability of software tools have enabled academics to employ e-assessment with less reliance on technology experts. Automated marking and handling, combined with production speed, allow for assessment on a greater scale and frequency than traditional assessment techniques, and promises to deliver more immediate and detailed feedback to learners, teachers and administrators (Ripley, Tafler, Ridgway, Harding, & Redif, 2009).

Measures can be taken to minimise cheating on computer based examinations.

Students taking computer-based examinations are likely to differ in attitude from candidates in traditional examinations. At least one study reported higher levels of (self-reported) cheating on computer-based tests; however, the form of cheating was a very traditional discussion between students who took the test at different times due to limited facilities (Butler, 2003).

The perception of fairness typically involves certainty of identification, and elimination of collaboration. In computer-based tests, identity assurance might be afforded by photographs (Rönnberg, 2001, cited

in Fluck, Pullen, & Harper, 2009), and collaboration is typically prevented by restricting computer operations to specific devices and network locations (Ko & Cheng, 2008, cited in Fluck et al.).

Suppressing collaboration, while necessary and normal, involves a reversal of habits carefully cultivated during the course. For example, in the course titled 'Information Processing and Technology' (IPT) in the NSW HSC, group work and long project management are topics and objectives in the syllabus, and are specifically assessed in Extended Tasks in Year 11. Theories of group work are explicitly addressed in the course. Communication and collaboration technology are areas of study in the course, and teachers seek to utilise them extensively to highlight their affordances. However, BOS policy requires that even group tasks be subject to individual assessment (Board of Studies NSW, 2003, s. 3.1.4). This is typically achieved by requiring students to clearly delegate and journal the activity in their group projects. Teachers seek to affirm student insights into roles and processes within the groups.

The falling cost and expanding capabilities of communication devices, such as mobile phones, combine in a growing challenge for invigilators. Communication through student-owned devices used in parallel with the testing equipment may be undetectable (Fluck, Pullen, & Harper, 2009, p. 512). Separate e-assessment technology which reduces computer functionality should be considered in those cases where Virtual Learning Environments (VLE) lacks effective security against cheating (Securexam, 2009; Gilbert et al., p. 29).

NAPLAN and other computer-based tests provided by BOS from 2002-2009 have been produced using Macromedia Director (Adobe, 2010). They run on the students' computer, but begin by ID and password up to a BOS server, and end by sending student responses and scores up to the BOS server, ensuring that a student can only take the test once. The test software stops the use of the computer for any other purpose while the test is running, to prevent improper communication during the test. Questions and answer pairs are compiled so they cannot be extracted from the software and leaked. Time-restrictions, extra-time and rest-breaks for selected students, coloured backgrounds for dyslexic students, and enlargements for visually impaired students are automatically applied. The school emulated some but not all of these capabilities, in the Quiz module in Moodle (Moodle.org, 2010).

Traditional, paper-based examinations in NSW are supervised by school staff to ensure that there is no collaboration or cheating. Care must be taken with the positioning of equipment, lest computer screens block the view of invigilators, and give students an opportunity to peek at the work of their neighbours. To minimise the opportunity for copying, the order of questions, and the sequence of options within questions, can be automatically randomised. Randomised questions must be designed to be disconnected from other content.

eAssessment offers good schools opportunities to reform school reports.

In NSW, completion of secondary school is marked by the Higher School Certificate (HSC). Students undertake a two-year program in 5-7 subjects. The syllabus for each subject is published by the state Board of Studies (BOS), but the instruction is developed and delivered by teachers in the school. In Year 12, about half of subject weight is assessed internally by the school, and the balance in a final examination set by BOS. The final certificate shows a mark based only on year 12 assessment. The summative assessment in year 11 is intended to also provide feedback for future learning (in year 12). In year 11, all assessment is designed and conducted internally by individual schools.

Schools are required by regulation (Board of Studies NSW, 2009) to provide detailed feedback to students, addressing the marking rubric, and advice for future learning, for each assessment task; and written reports to parents each semester, including a summary letter grade, and advice for future learning. Commonly, the semi-annual reporting takes the form of a page per subject, listing the tasks with scores, a general comment about attitude and progress, and a general recommendation for future learning. The comments are composed by class teachers, but expressed in the official voice of the school. Semi-annual student reports are official publications of the school. They normally

emphasise achievement rather than difficulty or failure, and do not show the school or teacher or student in a bad light.

Student reports are individual and confidential. Schools are required by regulation to indicate both each student's criterion-referenced mark and the average mark of her cohort, for comparison. As the final HSC marks will be aggregated to compute a rank position in relation to the whole state, parents are keenly interested in relative performance.

Schools are expected to provide opportunities for parents to discuss reports with the teachers (Cuttance & Stokes, 2000, p. 51). In many schools, teachers have duties which keep them on site after the students' last day, and they respond to parent contact then.

It is usual for students to drop one of their subjects after Year 11, in the light of their results. This allows a student to take on 'Extension' subjects which are only offered in Year 12, or to reduce her total workload, or to concentrate effort on a smaller number of subjects. Sometimes, from a small Year 11 class, withdrawal of a small number of students can make the course uneconomical for the school.

Course evaluations by students are not required by BOS. Some teachers make a point of soliciting direct comment through questionnaires, polls or learning journals. To permanently capture a consensus, the students may be encouraged to discuss the content and their experience of the subject in an online forum throughout the course.

There are opportunities for trials in a NSW HSC course

The IPT course spans several styles of content and a variety of assessment practices. Although it requires development of technical skills, it has a heavy emphasis on social impact analysis. The final (Year 12) examination requires students to analyse case studies and propose solutions to the technical and social challenges contained in them, in writing. Extended tasks earlier in Year 12 call for creative production and evaluation of technical solutions to specific practical scenarios. Assessment in the Year 11 course (Preliminary) is intended to confirm learning of the foundational knowledge, principles and techniques, and to develop production and evaluation skills.

Fairly detailed recommendations on assessment method and weighting have been provided by BOS for Preliminary IPT (Board of Studies, 2003). The school must set an individual project, a group project, at least one examination and one or two other tasks. All syllabus objectives must be assessed. Students must be provided with the rubric in advance and must receive written feedback on all tasks. Portfolios, checklists and profiles are not commonly used for formal assessment in this subject, although they are used at earlier year levels.

The pressure on schools to enable students to score highly in the external exam prompts teachers to speculatively predict future examination scenarios. The Digital Education Revolution initiative (DER) would be a suitable focus for a case study in 2010 or 2011. DER will equip all Year 9-12 students in Australia with internet-connected computers by 2012, lowering the implementation cost of online assessment in external, standardised examinations. An increase in online assessment is likely, but significant technical and social issues are involved. Teachers of IPT are therefore seeking to give IPT students firsthand experience of an online assessment, before starting Year 12.

The Board-approved syllabus specifies content and assessment methods, but not teaching and learning approaches. It is customary for schools to adapt the course to utilise local facilities and opportunities. The scale of the project-based assessment and teaching allows for significant modification of projects, and can give sufficient flexibility to accommodate students with exceptional needs.

Australian teaching practice, especially around the productive skills and group work, tends to be informed by Constructivist theory, although this is less apparent in digital work (Baker, 2009, p. 2). Assessment, on the other hand, has a standardisation imperative. To manage the tension, teachers tend to use recall or performance tasks for grading purposes, and divergent or creative tasks for learning activities. However, syllabus objectives, and hence, assessment criteria, are generally expressed as broadly observable actions. Typically, they fall far short of the detail required of indicators of learning. (See Figure 1.)

The primary purpose of examinations are usually summative assessment, reporting the knowledge and competence achieved by students. Students expect their examination mark to appear on their end-of-year report. IPT examinations focus on specific topics studied throughout the course, which have not been assessed in Extended Tasks earlier in the year.

Learning is not terminated by the commencement of a final examination. The individual feedback provided by or following e-Assessment should be exploited. Some practitioners assert that the terms *formative* and *summative* apply to the *use* of assessment data, not to the processes of collecting it (Gilbert et al., pp. 28, 32). Under any definition, an online task can automatically provide meaningful feedback for learning on every item.

Situated between the Preliminary and HSC year, a Year 11 examination is also formative assessment. Automated marking can enable students to review their test online and see their marks, model answers, and some informative guidance for future learning within 24 hours of completing it. An automated feedback system can be loaded with answer scores, correct answers and suggestions for improving learning or performance, such as: "Read Chapter 3, and discuss question 3.11."

Online testing technology does need to be proven in the technical context of the school, and this, too, can be automated. The performance of the technology can be logged, and reviewed by faculty staff along with evaluation comments by students. De-identified extracts from technical logs can subsequently be analysed by staff and students.

In line with the requirements of the Board of Studies, online examinations are naturally criterion-referenced tests. The questions can be related to syllabus dot-points in order to demonstrate achievement of specific syllabus objectives.

In the HSC, there has been a trend toward increasing the weight of Objective Response questions, and to atomisation of the exam-marking process. Examples include consideration of Objective Response questions in English (Board of Studies NSW, 2008), and the fragmented marking of Mathematics papers (Taylor, 2009, p. 24). A test can be designed to utilise Objective Response questions and provide a numerical total score. In compliance with BOS requirements, the weight of each question can be shown on screen during the test. The combination of these measures can make the arithmetic basis of the final score explicit to students. Numerical scores can be converted to and reported alongside a letter grade. (The boundary scores for each grade should be published to students.)

Conclusion

Considering the variety of forms and mutability of computer-based assessment, it is difficult to talk about e-assessment as a single enterprise. However, typical characteristics and effects have been observed and can be exploited to achieve more reliable measurement or more useful formative feedback. eAssessment offers good schools opportunities to reform school reporting schedules and content. A new abundance of equipment is creating an opportunity for computer-based testing, which should be seized, with appropriate measures to minimise cheating. This paper has outlined some considerations for a trial of e-assessment within an IPT course at Year 11. A trial along these lines has been made, and observations and reflections will be discussed in a subsequent paper.

Bibliography

Adobe. 2010. *Macromedia Director 11.5*. Retrieved 06 06, 2010, from Adobe.com: http://www.adobe.com/products/director/

Baker, R. 2009. *Pedagogies and Digital Content in the Australian School Sector.* Education Services Australia.

BC Consulting. 2006. Effective on-line assessment: Workshop support materials. Sydney: University of New South Wales.

Beaton, A. E., & Zwick, R. 1990. *The effect of changes in the National Assessment: disentangling the NAEP 1985-86 reading anomaly.* Princeton, NJ: Educational Testing Service.

Board of Studies. 2003. *Information Processes and Technology Stage 6 Syllabus.* Sydney: Board of Studies NSW.

Board of Studies NSW. 2003. *HSC assessment in a standards-referenced framework - A Guide to Best Practice.* Board of Studies NSW. < http://www.boardofstudies.nsw.edu.au/manuals/hsc_assessment_std_ref_fwk.html> [Accessed 6/06/2010]

Board of Studies NSW. 2009, June 19. *New HSC Assessment Support Document*. Board of Studies. <http://www.boardofstudies.nsw.edu.au/syllabus_hsc/newhsc_assessment.html> [Accessed 6/06/2010]

Board of Studies NSW. 2008. NSW Board of Studies Proposed Changes to HSC examinations and school assessment. Sydney: Board of Studies NSW.

Bransford, J. D., & Franks, J. J. 1976. The role of "effort after meaning" and "click of comprehension" in recall of sentences. Educational Resources Document Reproduction Service (ERIC).

Butler, D. 2003. Impact of computer based testing on student attitudes and behaviour. *The Technology Source* .

Byrnes, R., & Ellis, A. 2006. The prevalence and characteristics of online assessment in Australian universities. *Australasian Journal of educational Technology , 22* (1), 104-125.

Cizek, G. J. 1991. The effect of altering the position of options in a multiple-choice examination. *Annual Meeting of the National Council on Measurement in Education.* Chicago, Il: Educational Resources Docuemnt Reproduction Service (ERIC).

Clariana, R., & Wallace, P. 2002. Paper-based versus computer-based assessment: key factors associated with the test mode effect. *British Journal of Educational Technology , 33* (5), 593-602.

Cuttance, P., & Stokes, S. 2000. *Reporting on student and school achievement.* Canberra: DETYA.

Fluck, A., Pullen, D., & Harper, C. 2009. Case study of a computer based examination system. *Australasian Journal of Educational Technology , 25* (4), 509-523.

Gardiner, J., Holmes, B., & Leitch, R. 2009. *Assessment and Social Justice.* Bristol, UK: Futurelab.

Gilbert, L., Gale, V., Warburton, B., & Wills, G. 2009. *Report on Summative E-Assessment Quality (REAQ).* Southampton: Joint Information Systems Committee (JISC).

Haas, C., & Hayes, J. R. 1986. What did I just say? Reading problems in writing with the machine. *Research in the Teaching of English , 20* (1), 22-35.

Hattie, J. 2010. *E-AsTTle home*. (Ministry of Education, Wellington, New Zealand) Te Kete Ipurangi The Online Learning Centre. <http://e-asttle.tki.org.nz/> [Accessed 6/06/2010]

Hogan, T. P. 2007. *Educational Assessment: A Practical Introduction.* Hoboken, NJ: John Wiley & Sons.

Ko, C., & Cheng, C. 2008. Flexible and secure computer-based assessment using a single zip disk. *Computers and Education , 50* (3), 915-926.

Linacre, J. M. 2003. What is Item Response Theory, IRT? A Tentative Taxonomy. *Rasch Measurement Transactions , 17* (2), 926-927.

MCEEDYA. 2008. *National Assessment Program - ICT Literacy Years 6 & 10 report 2008.* Carlton: MCEEDYA.

Moodle.org. 2010. *About Moodle*. < http://docs.moodle.org/en/About_Moodle> [Accessed 1/05/2010]

Mourant, R. R., Lakshamanan, R., & Chantadisai, R. 1981. Visual fatigue and cathode ray tube display terminals. *Human Factors , 23* (5), 529-540.

Partchev, I. 2004. *A visual guide to item response theory.* Germany: Friedrich-Schiller-Universität Jena.

Ripley, M. 2007. *E-assessment - an update on research, policy and practice.* Bristol: futurelab.

Ripley, M., Tafler, J., Ridgway, J., Harding, R., & Redif, H. 2009. *JISC Final Report - Review of Advanced e-Assessment Techniques.* Martin Ripley Ltd. JISC.

Rönnberg, K. 2001. Security in learning management systems - User authentication in online assessment. *Master's Thesis.* Sweden: Umea University.

Securexam. 2009. *Securexam - Frequent Questions*. <http://www.softwaresecure.com/faq.htm> [Accessed 6/06/2010]

Taylor, C. 2009. *Annual Report 2009.* Sydney: Board of Studies.

Wilson, R. F. 2001. *HTML E-mail: text font readability study. Results of a survey conducted April, 2001.* <http://www.wilsonweb.com/wmt6/html-email-fonts.htm> [Accessed 6/06/2010]