Screen-Based Assessment

Stephen J Swithenby
Centre for Open Learning
of Mathematics,
Science, Computing and
Technology
The Open University
Milton Keynes   MK14 6AT
s.j.swithenby@open.ac.uk

A major driver for the adoption of screen-based assessment has been the increased intervention of Government and its agencies in promoting teaching, learning and, more specifically, e-learning

Abstract

Inexorably and across several fronts, screen-based assessment is becoming a major part of the experience of university students, particularly but not exclusively in the sciences. This movement reflects the emphasis the Qualifications and Curriculum Authority (QCA) is giving to the development of screen-based assessment at secondary level, where the universal availability of an e-assessment option in high stakes exams is an adopted goal[1].

The drivers for this change are economic, pedagogic and opportunistic. Rapid technological progress is facilitating the wider availability of computer based tasks that reflect authentically the learning outcomes of science courses. There is growing experience in the design of such tasks, with increasing commercial involvement, particularly in the USA. An examination of theories of assessment demonstrates that there are sound pedagogic reasons to pursue these developments.

The main focus of this review will be assessment for which a computer acts as a means of delivery, grading and feedback. I will outline the capabilities of contemporary systems, illustrate some good practice, and identify areas where the use of the technology is moving forward rapidly. There are exciting developments in the grading of free format responses, in diagram or text form, which are now emerging on a pilot basis. Of particular interest is the assessment of higher order cognitive and subject skills. Also important is the potential for item banks that can allow the sharing of the costs of authorship. Several of these issues are reviewed more fully in Conole and Warbuton[2].

Finally, I will comment briefly on assessment that is facilitated by computers without the computer acting as a grading tool. At a mundane level, this might involve the electronic submission of traditional assignments. Of more interest are electronically mediated peer assessment, the generation of e-portfolios, the grading of screen based experimentation and the evaluation of the student's performance in contributing to computer based group activities, eg Wikis, electronic conferences, etc.

Drivers for e-assessment?

E-assessment can offer increases in efficiency, though, clearly, the advantage that is realised depends on the scale of use of any assessment that is generated. Within the pedagogic model of the lone teacher and their class, efficiency gains are difficult to realise in all but the least complex assessments implemented under standard Virtual Learning Environments (VLEs). This explains the preponderance of basic 'progress check' quizzes, most of which are content based and used as a means of securing the engagement that is a prerequisite of learning[3,4].

There is limited rigorous data on the efficiency gains achieved by university e-assessment processes though it has been noted that the promise has not been translated quickly into practice[5]. A recent case-study based investigation has suggested that, in some prominent initiatives, pedagogical need has been the dominant driver for introduction of e-assessment[6].

It might be argued that e-assessment is a necessary part of a contemporary pedagogic strategy in that learning experiences that are increasingly mediated through screen activities should be assessed using similar media. However, the culture of regarding an *unseen* pen-and-paper exam as the 'gold standard' is still strong, in spite of the rapid advance of screen-based high-stakes assessment at secondary levels and in skills-testing by, for example, the UK Driving Standards Agency with its national network of testing venues.

Over the last few years, there has been a rapid growth in understanding of the formative role of e-assessment. The potential is evident when e-assessment capabilities are examined alongside criteria for assessment to produce learning, eg that devised by Gibbs and Simpson[3] using empirical and theoretical arguments. Key issues in their analysis are the regularity and quality of student engagement, the timeliness and quality of feedback and the student engagement with that feedback. Computer based assessment is completely flexible in its time of delivery and can produce high levels of engagement. The feedback from e-assessment systems is available instantly and can be differentiated with some sophistication (eg Rayne et al[7], Jordon et al[8]). Feedback loops can be closed by directing the student to further questions or other scheduled activities.

A major driver for the adoption of screen-based assessment has been the increased intervention of Government and its agencies in promoting teaching, learning and, more specifically, e-learning. This is evident in the targeted funds associated with the Fund for the Development of Teaching and Learning (FDTL) projects, the many Joint Information Systems Committee (JISC) initiatives and the Centres for Excellence in Teaching and Learning (CETL). Examples of activities generated via these funds are described below.

To an increasing extent, large publishers are becoming aware of an assessment market that is complementary to the conventional textbook sector. It has become commonplace to receive e-resources with the purchased book. Such resources might include an e-book, animations and simulations and end-of-chapter tests. This assessment element is expanding in volume and sophistication of construction. For example, John Wiley and Sons is now offering *Wiley PLUS*, a package that includes assignments that can be constructed from questions that are organised by chapter, level of difficulty, and source, and which include feedback . Students' responses are automatically graded, and the results recorded in a gradebook. With ~20 *Wiley PLUS* packages available in the Physical Sciences, including several standard texts, and compatibility with the market leading *Blackboard/WebCT* environment, it is likely that academics will find it attractive to adopt the embedded computer aided assessment, at least for non-summative purposes.

The summative role of screen based or e-assessment has been limited by many factors. There is the perception that machine generated tasks are too closed to represent authentically the full range of learning outcomes of a given programme. So, for example, one might test whether a student is able to recall the steps in a chemical synthesis or solve a problem in electromagnetism but might encounter more difficulty in assessing learning outcomes involving group work, communication or creativity. Anecdotally, this view has been challenged in a series of recent workshops in which academics were able to devise relevant though not always sufficient tasks for all the learning outcomes of an illustrative inter-disciplinary science course. Other anxieties concern; collusion, plagiarism, recognition of partial achievement, the logistical problems of engaging simultaneously an entire cohort, and, by no means least, institutional policies. Warburton[9] has analysed such obstacles within a framework for the introduction of computer aided assessment that contrasts a gradual low risk strategy, through quizzes and progress checks, with one that involves summative

assessment and is high risk. He suggests that adoption of high risk strategies with consequent frequent failure has offered ammunition to those who oppose computer based assessment for cultural reasons.

Conservatism in the use of summative computer based assessment is not universal. From 2007, the Medical College Admission Test, which is a required part of any application to a medical school in the USA, will only be available online and will be administered via testing sites operated by Thomson Prometric. Illustrative questions are available via the Website (http://www.aamc.org/students/mcat/start.htm). The total test takes 5½ hours and includes sessions on both Physical and Biological Sciences. Most questions are simple in construction and require recall or interpretation. The advent of large commercial concerns offering 'total assessment solutions on a global scale' to professional bodies, Government agencies, universities and companies is an indicator of a trend which will undoubtedly affect universities.

Contemporary Systems
Systems that can be used by teachers to create a computer based assessment are now widely available. VLE systems such as *Blackboard/WebCT* and *Moodle* have limited intrinsic capabilities and, to date, the most developed computer assisted assessment (CAA) systems have been constructed using specialist packages, eg *Questionmark Perception*, *Tripartite Interactive Assessment Delivery System (*TRIADS), *Maple TA*, *SoftwareTeaching of Modular Physics (*SToMP), and *Open Mark*.

Questionmark Perception has been adapted widely (eg Ellis et al[10]). The company lists 21 question types that are supported by the software though many of these can be regarded as technical variants on the same fundamental task. For example there is no fundamental difference between a multiple choice question and one that is based on a drag and drop operation. A useful facility provide by the package is an interface that allows incorporation of *Java* or *Flash* elements. A relevant illustration of the use of this platform is provided by the Computer Aided Assessment in Chemistry project whose outputs are available online[11]. The demonstration of *Questionmark Perception* question types available via http://www.science.ulster.ac.uk/caa/index.html/ provides a useful introduction. Of particular interest are laboratory preparation questions that allow the student to construct a virtual experiment, thereby involving themselves in a learning task that transcends the superficial requirement for recall.

The *Tripartite Interactive Assessment Delivery System (*TRIADS) was originally developed via an FDTL project. It is now hosted by the Centre for Interactive Assessment Development (CIAD) at the University of Derby. It too provides for author flexibility through multiple question types, feedback facilities etc. The *Online Assessment and Feedback (*OLAAF) project[12] has provided a powerful example of the application of the TRIADS system. The tactic employed by the OLAAF authors has been to develop 'relatively few test items (not large item banks!) to detect critical misconceptions that may impede learning.' Feedback, crafted with equal care is then used to prompt appropriate further learning. The tests developed to complement a first year Molecular Biology module show how computer based assessment can be devised to cover learning outcomes at all levels with task and feedback that facilitate learning.

The authors have succeeded in devising questions that involve the students in the construction of knowledge.

Maple TA, although aimed at mathematicians and based around the *Maple* mathematics engine, is of relevance to physical scientists. It has been exploited elegantly by Greenhow and colleagues[13] to provide mathematics questions that, through the use of variables in the authorship, can be instantiated with random (within physical limits) values to yield endlessly repeated variants and appropriate feedback[10].

OpenMark allows three feedback stages and flexible grading plus, in common with the other systems, is a powerful pedagogic tool. A major constraint in its use is the complexity of authorship. In this case, XML code is required and, although based around templates, the most practical mechanism for realisation is collaboration between an academic who has the subject expertise and a learning technologist/software engineer. This problem exists for all systems at both the pedagogic and technical levels. Warburton[9] has highlighted the problem. There is a tension
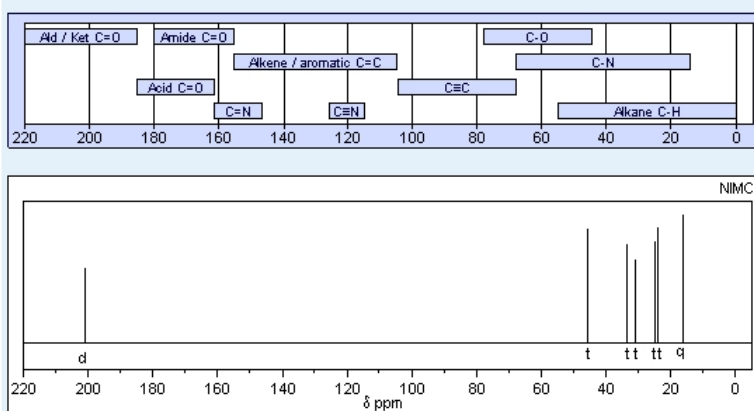
Figure 1: Java applets can provide additional (dynamic) functionality within assessment packages. In this example, a molecular editor is used to allow student construction of a response molecule.

OpenMark[14] is a system developed by the Open University that is used for both formative and summative online assessment. An online demonstration site http:// www.open.ac.uk/openmarkexamples shows *OpenMark* supporting a range of responses;

● Text responses (eg simple text entry, chemical formulae, mathematical formulae and structured responses)
● Numeric responses (eg single entry, multiple entries, use of scientific notation, evaluation of significant figures and units)
● Multiple choice responses (eg including single choice, multiple choice, drag and drop, words within text, drag and drop words onto images, drag and drop images)
● 2d responses (eg placing a marker, drawing a line and the use of a *Java* molecular editor)

Most of these facilities are available in other systems. An exception is the use of a *Java Molecular Editor* (Figure 1), an open source tool devised by Novartis.

between the need for technical and pedagogical design skills in the design of advanced computer aided assessment and the traditional academic ownership of teaching in UK Higher Education. This problem is less evident in the newer universities who tend to have greater central control over the teaching process.

Two ways of potentially overcoming the expertise problem are the use of authoring tools and item banks.

As well as the authoring tools available in commercial systems (eg *Questionmark Perception* includes 'wizards' for all of its question types), there has been a recent major attempt to provide a generic solution. The JISC-funded *Technologies for Online Interoperable Assessment (*TOIA) initiative aimed to 'remove many of the barriers for teachers who wish to move into computer-assisted assessment - and avoid lock-in to a particular proprietary system.' Although not widely adopted, its use of Web based templates that are structured so as to provide the information necessary to work within the interoperability framework provided by the IMS Question and Test Interoperability (QTI) specification may provide a route that wins eventual acceptance.

The QTI specification provides a well developed and comprehensive mechanism for interoperability of computer based assessment systems. It includes a large range of question types, and specifies the metadata required for the questions to be incorporated within assignments and for reporting of results. Although several systems claim QTI compliance, the capability is fragile in most cases and import of questions from a different package may not be a practical option for the non-technical user. It is not clear whether QTI compliance will increase as the advantage of compliance is of limited value to vendors.

There are now several item banks available in the UK in the physical sciences. The largest and most accessible resource is provided by commercial text book publishers. The FDTL initiative has spawned a number of contributions although few

The Higher Education Academy Subject Centre for the Physical Sciences is a gateway to several collections created through development projects[11,15]. These are implemented on a variety of platforms and are mainly in Chemistry. Direct contact with the author(s) may facilitate their use. Recent initiatives are increasing the breadth of coverage and a QTI compliant question bank is now under construction.

The interoperability issues that hinder use of question banks may be finessed by their availability as a Web service. In this case, the interoperability issue is addressed through standard mechanisms, eg use of XML. This technique underlies the commercial assessment packages that are widely used in the USA. A leading example of such a system is the *Mastering* Physics 'homework and tutorial' package (Figure 2) marketed by Pearson Education. This subscription funded Web

Figure 2: A screenshot from Mastering Physics.

with core relevance within the physical sciences. One of the most fully developed is the *Electrical and Electronic Engineering Assessment Network (*E3AN) which developed a database of nearly 1400 peer reviewed questions involving knowledge, comprehension, application and analysis.

resource includes learning materials, assessment and grading. In effect it functions as a discipline specific VLE. The company claim to have graded and provided feedback on 50 million student submissions of *Mastering Physics* assignments to date. All that is needed to access the service is a Web browser and a credit card!

A crucial issue in determining the viability of such Web services is the expectation of the customers. In the USA, students pay to access the systems determined by their teachers. Whether this expectation is exportable to a UK society which has enjoyed free higher education until recently is unclear.

Future developments
To date, virtually all computer based assessment has involved tasks with heavily circumscribed input mechanisms, eg ticking of boxes, dragging and dropping objects, manipulating graphs, entering words, numbers, expressions or equations, etc. There is a stark contrast between such activities and the writing of an extended essay or report. Given that there is a strong belief that such open response formats are needed to assess higher order learning outcomes, and the demands of marking and providing feedback on essays is so high, it is reasonable to ask whether computers can offer accurate simulations of the human teacher. Perhaps surprisingly, some progress has been made[16,17]. These studies are limited to short free text answers, eg up to five lines, and use variants on an information extraction technique based on the construction of answer templates generated by experts. Mitchell et al[18] have achieved impressive results in a medical student progress check that uses a database of 270 (very) short answer questions. After a period of intensive development the rate of disagreement between an expert moderator and the computer based system was reduced to 0.6%, which compares favourably with the ~5% disagreement between human markers in comparable studies.

Marking of longer texts is less developed though there are well established tools for automatic feedback on writing style: content, grammar, usage, style, organisation etc. (eg ETS[19]). Such systems tend to rely on a databank of expert-marked assignments with which submitted assignments can be compared. Further work in this area and in the complementary problem of marking free diagrams is ongoing and progress may be expected.

The above discussion has centred on assessment focused tasks. Although formative, the assessment function is at the core of the design of the experience. However, there is a growing wealth of computer based activities that are designed as constructivist learning experiences. Examples are; simulations and virtual practice environments, group e-working within conferences or in Wiki construction, e-portfolio generation, etc.

The QTI specification provides a well developed and comprehensive mechanism for interoperability of computer based assessment systems. It includes a large range of question types, and specifies the metadata required...

In each case, an integrated assessment strategy would allow the computer activity to serve as input to an automatic grading and feedback tool. Strategies for such approaches are only now beginning to emerge with an accent on monitoring the quantity and timing of activity rather than the quality of the contribution. So, for example, it is straightforward to monitor when a student is making an input to a discussion forum – it is less easy to identify the value of that contribution.

One pedagogic strategy that has attracted significant interest recently is peer assessment. This has the merit of engaging the student with the performance criteria and can engender a feeling of community responsibility. Several workers have demonstrated how such strategies might be facilitated by the user of computers (see for example Davies[20]).

The strongest message from this review must be the need for collaboration in developing screen based assessment. There is little evidence that the pedagogic skills needed to develop more sophisticated computer based assessment materials are common in universities and there is limited experience of the newer integrated assessment strategies. Given the many pressures on academics, it is unlikely that the skills base will expand quickly and successful development strategies are likely to involve pedagogists, subject experts and software engineers collaborating to produce resources that can be shared. Vendors have adopted such an approach and are now offering high quality learning and assessment resources.

References

1. Qualifications and Curriculum Authority A proposed blueprint for e-assessment (2004). Available online from http://www.qca.org.uk/downloads/6995_blueprint_for_e-assessment.rtf. (Accessed 7/8/06)
2. Conole, G. and Warburton, W., A review of computer-assisted assessment. ALT-J, **13**(1), 17 – 31 (2005)
3. Gibbs, G. and Simpson, C., Does your assessment support your students' learning? *Learning and Teaching in Higher Education* (on-line), **1**(1), 3-31 (2004-5). Available online at http://www.glos.ac.uk/adu/clt/lathe/issue1/index.cfm (accessed 7/8/06)
4. Nicol, D. and Macfarlane-Dick, D., Rethinking formative assessment in HE: a theoretical model and seven principles of good feedback practice (2004). Available online at http://www.heacademy.ac.uk/resources.asp?process=full_record&section=generic&id=353 (accessed 23rd October 2005)
5. Conole, G. and Dyke, M., What are the affordances of information and communication technologies? ALT-J, **12** (2), 113 - 124 (2004)
6. Whitelock, D. and Brasher, A., Developing a roadmap for e-assessment: which way now? (2006). JISC report 2006 - to be published
7. Rayne, R. and Baggott, G., Computer-based and Computer-assisted Tests to Assess Procedural and Conceptual Knowledge (2004). Proceedings of the 2004 CAA Conference (Loughborough UK). Available online at http://www.caaconference.com/pastConferences/2004/proceedings/index.asp (Accessed 13/11/06)
8. Jordan, S. and Swithenby, S. J., On-line summative assessment with feedback as an aid to effective learning at a distance. Proc of 2004 ISL Symposium Diversity and Inclusivity (Ed C. Rust) Alden Press, pp 480-485 (2005)
9. Warburton, W., Quick win or slow burn? Modelling UK HE CAA uptake (2006). To be published in the Proceedings of the 2006 CAA Conference (Loughborough UK)
10. Ellis, B. N., Gill, E., and Greenhow, M., Recent Developments in setting and using Objective Tests in Mathematics using QM Perception (2005). Proceedings of the 2005 CAA Conference (Loughborough UK). Available online at http://www.caaconference.com/pastConferences/2005/proceedings/index.asp (Accessed 7/8/06)
11. Adams, K., Byers, W., Cole, R., Ruddick, D. and Adams, D., Computer-aided Assessment in Chemistry. Development Project Report. UK Higher Education Academy Physical Sciences Centre (2004). Available on online at http://www.physsci.heacademy.ac.uk/Resources/DevelopmentProjectsReport.aspx?id=78 (Accessed 7/8/06)
12. Online Assessment and Feedback OLAAF (2004) Access to the resources can be solicited via the Website contacts. See http://www.bbk.ac.uk/olaaf/index.html
13. Greenhow, M., (2006). Details of the CAA systems developed by Martin Greenhow and colleagues are available via http://people.brunel.ac.uk/~mastmmg/ (accessed 7/8/06)
14. OpenMark (2006). A demonstration site is at http://www.open.ac.uk/openmarkexamples
15. Price, G., Computer Aided Assessment and Feedback. Development Project Report. UK Higher Education Academy Physical Sciences Centre (2006). Available online at http://www.physsci.heacademy.ac.uk/Resources/DevelopmentProjectsReport.aspx?id=213 (Accessed 7/8/06)
16. Pulman, S. G. and Sukkarieh, J. Z., Automatic Short Answer Marking, in Association for Computational Linguistics: Proceedings of Second Workshop on Building Educational Applications Using NLP, Ann Arbor, Michigan, pp 9-16 (2005)
17. Mitchell, T., Russell, T., Broomhead, T. P. and Aldridge, N., Computerized marking of short answer free-text responses (2003). Paper presented at the 29th Annual Conference of International Association for Educational Assessment IAEA, Manchester UK
18. Mitchell, T. Aldridge, N., Williamson, W. and Broomhead, P., Computer Based Testing of Medical Knowledge (2003). Proceedings of the 2003 CAA Conference (Loughborough UK). Available online at http://www.caaconference.com/pastConferences/2003/procedings/index.asp (Accessed 7/8/06)
19. Educational Testing Services ETS at http://www.ets.org. The e-rater tool is relevant
20. Davies, P. Don't Write, Just Mark; The Validity of Assessing Student Ability via their Computerized Peer-Marking of an Essay rather than their Creation of an Essay. ALT-J (CALT) **12**(3), 263-279 (2004)