European Educational Research Journal, Volume 5, Number 2, 2006 **DOI: 10.2304/eerj.2006.5.2.122**

Use of Online Assessment to Enhance Teaching and Learning: the PASS-IT Project

HELEN ASHTON
Heriot-Watt University, Edinburgh, United Kingdom
CHRISTINE WOOD
Scottish Qualifications Authority, United Kingdom

ABSTRACT This article describes a recent collaborative project (PASS-IT) which investigated the use of online assessment in secondary and further education in Scotland. The aim of PASS-IT was to explore the potential of formative and summative assessment in secondary education, and to build on previous research into the applicability and validity of online assessments. An overview of the project is given. Examples of online questions are provided in section 2 with a discussion of some of the issues and benefits. Section 3 outlines the main research findings of the project, which provide evidence for the validity of online assessments.

Introduction

It is widely acknowledged that assessment has a significant impact on the learning process. As well as grading for an award, assessments can diagnose strengths and weaknesses, measure progress, and provide useful feedback to both teachers and students in order to develop an appropriate learning plan. Technology is playing an increasing role in everyday life. In education, it is important to consider the issues of using computers in the measurement of learning.

PASS-IT (Project on ASsessment in Scotland – using Information Technology, http://www.pass-it.org.uk) was a 27-month project (August 2002-December 2004) which involved many of the main education bodies in Scotland in secondary education: the Scottish Qualifications Authority (SQA), Learning and Teaching Scotland (LT Scotland), the Scottish Further Education Unit (SFEU), the Scottish Centre for Research into Online Learning and Assessment (SCROLLA), and BBC Scotland.

In Scottish secondary schools and colleges, the National Qualifications (Access, Intermediate, Higher and Advanced Higher) are taught in three units. Each of these units is internally assessed, with a pass in all three units enabling the student to sit the externally assessed examination at the end of the course. The internal assessments (usually referred to as the NABs – National Assessment Bank) are set to measure minimum competencies and are pass/fail only. The awarding body (the SQA) provides a number of versions of the NAB tests, which the teacher administers when they feel the students are prepared. The classroom teacher does the marking, and reports the results to the SQA who record and moderate the process. It is the online provision of these NAB unit tests that was investigated during PASS-IT.

From the outset, the intention was to provide both formative and summative online assessments. The purpose of the formative assessments was to enable students to become comfortable with the system and to allow formative assessment of their learning. These assessments would be available prior to students sitting a summative examination. PASS-IT aimed to:

- develop, pilot and evaluate online assessments in a range of subjects and levels;
- investigate how online assessments could support teaching and learning;

- explore the possibilities of helping candidates with additional needs;
- examine quality assurance issues; and
- construct a model for the process of development, good practice and moderation of the NAB tests.

This illustrates the main aims, although other issues such as interoperability and accessibility also formed a significant part of the work.

Prior to PASS-IT the research had concentrated on the use of online assessment in Higher Mathematics (Fiddes et al, 2002; McGuire et al, 2002). This work had utilised question types beyond that of multiple choice and it was considered important to investigate the appropriateness of different question types over a variety of levels and a range of subjects. The following subjects and levels were included in PASS-IT:

National Qualification Units

- Mathematics: Access 2 & 3, Intermediate 1 & 2, Higher, Advanced Higher
- Chemistry: Higher
- French: Higher (Reading/Listening)
- Music: Higher (Listening)
- English: Intermediate 1

Higher National Certificate Units

- Computing

Before discussing the research aims and results, it is worth pausing to consider the kinds of questions that are being posed by the online assessment system. This is valuable from the point of view of familiarisation with the system being used, but also as it highlights some of the useful features, challenges and benefits that have been discovered.

An Overview of the Online Assessments

The main components of the PASS-IT assessment system are the authoring tools, the delivery engine, and the student and teacher reporting systems. The online reporting systems are not discussed here, but descriptions of the reporting system and discussions on the importance of reporting in online assessment can be found in the articles by Schofield & Ashton (2005) and Ashton et al (2004).

The online assessments are served through a conventional web browser by the delivery engine. The candidate views a single question at a time. Each question may have several keyparts requiring an answer, and may have optional steps (see Ashton et al [2003] for a description of steps). Each keypart (or step) can take one of the following types: Information only, Multiple Choice; Multiple Response; Hotspot; Wordmatch; Essay; Judged Mathematical Expression (JME); and Integrated Multimedia (Ashton et al, 2003; Thomas et al, 2004). Examples of some of these part types can be seen later in this article.

A typical question can be seen in Figure 1. This question comes from Mathematics at the level of Intermediate 2. The student is presented with a statement of the problem and the associated graphic (collectively referred to as the *Main Question Text*). The keyparts appear below this (labelled 8.1 in this case) with a box for the student to enter the answer and a submit button.

When the students are happy with their answer they press the submit button, and this commits the answer to the database at the remote server. Following a successful submission of an answer, the submitted answer is mirrored back to the student in the 'Your currently accepted answer:' field beneath the input box. This is crucial in formative assessment to allow immediate reflection, and in summative mode, to enable the student to check the answer has been recorded and to enable the question to be revisited and the answer modified if desired. In the question illustrated in Figure 1, keypart 8.1 requires a mathematical expression answer (or a Judged Mathematical Expression – JME). There are two features which are specific to this part type and have proved valuable during the project: rendered answers and mathematical equivalence.

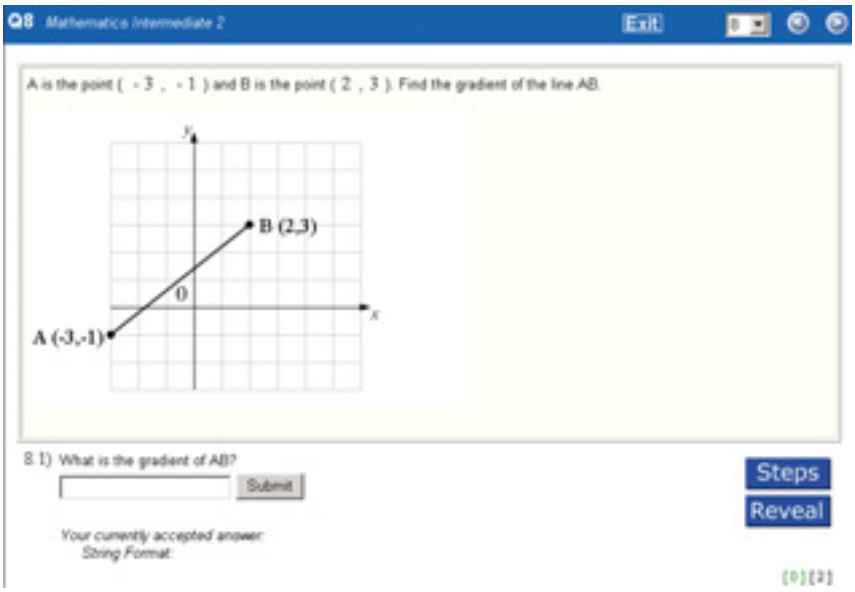

Figure 1. A question from Intermediate 2 Mathematics.

Rendered Answers

As has been mentioned, all question parts have the accepted answer mirrored back to the student, but for a Mathematical (JME) part type, this is mirrored back in two ways:

- the string format (similar in nature to the format used on many calculators), for example 'x^2', and
- the rendered format, x^2.

This reassures the candidate as they can see how the assessment engine has interpreted their mathematical answer. Many users saw this as a very positive feature following its implementation in the second half of the project.

Mathematical Equivalence

The question in Figure 1 has the correct answer of 4/5; however, a candidate may wish to express this in a different format, for example as 0.8. Both are mathematically equivalent. The PASS-IT assessment system marks mathematically equivalent answers as correct (although it only requires the question author to specify one format of the answer). In addition, the system can disallow certain formats of mathematically equivalent answers using a variety of techniques that are specified at authoring time such as 'not allowed' or 'must have' strings of characters. This approach provides a convenient and reliable method of marking mathematically equivalent answers without placing an additional burden on the question author.

This question (Figure 1) also illustrates several other features that proved invaluable in this project in both supporting formative assessment prior to summative assessments, and to support the increasing demands for flexibility and extending question types.

Randomisation (or Creating Templates)

A feature worth noting here is that of randomisation (NB: this is not restricted to mathematical part types). Questions can be authored which use random parameters that enable the reuse of a question. For the question in Figure 1 this means that the coordinates of the points A and B can be randomly generated by the assessment system, enabling the question to be used on multiple occasions for repeated practice and reinforcement – something which has proved to be valued by

both student and teacher. These random parameters can be used for the display of information, to describe the correct answer, and can be employed in the accompanying graphic. It is features of this kind that enable the question author to create useful, educationally valid questions that can support learning and development.

Integrated Multimedia

Whilst random parameters are not a new feature of this system, the need for wider use of other multimedia components became important during PASS-IT. This requirement came from the desire not just to randomise parameters, but also to randomise the accompanying graphic, especially where it was desirable for the candidate to learn to interpret the graphical information. In the case shown in Figure 1 this means that the coordinates and the straight line represented in the graphic are appropriate for the specific instance of the randomised problem.

As well as presenting randomised graphics, the need to extend the answering mechanisms became more crucial. An example of this can be seen for Music in Figure 2.

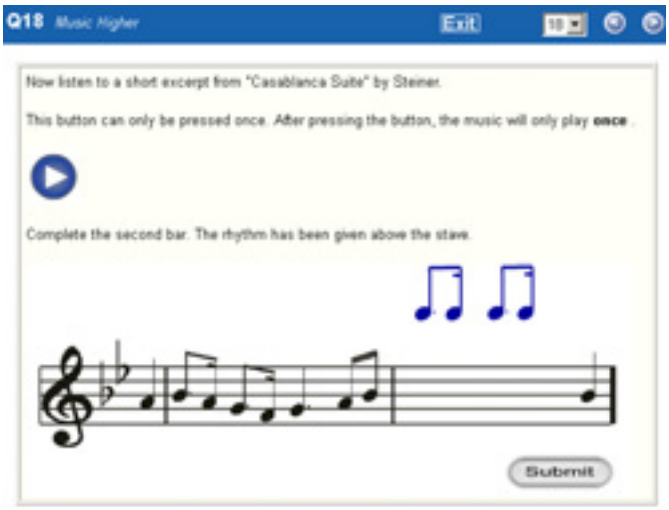

Figure 2: A Music (Higher) question showing the use of interactive multimedia.

Figure 2 shows an example of interactive multimedia as an answering mechanism, where the candidate listens to a piece of music and places the musical notes on the stave. In the previous example of a randomised graphic, which appears as a static graphic to the user (Figure 1), the communication was from the assessment engine to the multimedia. In this example the multimedia is used not only for presentation, but also as an answering mechanism. In this case the communication is in the opposite direction, with the embedded multimedia communicating the answer (the position of each bar) to the assessment engine. In addition, this mechanism was used to ensure the candidate could only listen to the music once.

Some candidates felt that it was easier to give a clearer answer using this mechanism (eliminating erasing or scoring out), and questions like these received positive comments from the candidates, such as:

'They were very good to answer as you only had to place the answer where you wanted it to go.'

'I liked these as they were easy to use and adjust.'

The interested reader is directed to Thomas & Ashton (2005) where further information can be found about the use, reuse and integration between the assessment system and multimedia components such as Macromedia Flash™ and JAVA.

Other features worth noting are the ability to deliver an assessment in different modes (Exam: with no feedback, Practice: with marks and ticks and crosses, and Help: Practice mode with the addition of a reveal answer facility), the facility to monitor and support formative usage, and the

attention paid to usability and accessibility. These and many other features of the system, alongside the ability to develop the system throughout the project to fulfil the requirements demanded by the various subjects and levels, enabled learning technologists and subject specialists to design questions which they felt would enable assessment of the appropriate learning objectives.

Further understanding of the assessment system can be obtained through the use of the PASS-IT assessment engine by exploring sample questions from the project. To do this please visit http://www.calm.hw.ac.uk/pass-it.html

An Overview of Some of the Research Issues

However, having a system that enables the authoring of a set of appropriate questions does not mean that online assessment should be utilised. Traditionally all the assessments under consideration in this project were taken using a pen and paper test (with additional audio for Music and French). In making the move to online assessments it is important that standards are maintained, and that the issues associated with developing questions for online assessment are understood. There are also issues of comparability between paper and computer-based assessments, as it is unlikely that all students across Scotland will start sitting online assessments at the same time, so paper and electronic tests are likely to sit side by side for a number of transitional years. In addition, this form of assessment is new to qualifications bodies, teachers and students, so it is important that there is confidence that the online assessments are appropriate, fair and valid.

Research into the issues of appropriateness and validity of online assessments was a crucial part of this project. Some of the areas investigated are discussed below.

The Issue of Equivalence

The starting point for assessing equivalence in this case was to ensure that the paper and computer formats of the NAB assessments provided an equal opportunity for students to demonstrate their level of ability in the specified learning points, and hence in the overarching learning outcomes. It was also important to demonstrate that the computer could measure the appropriate learning points identified by SQA for their tests.

Since all of the assessments existed in paper format, the approach taken was to transform the paper-based questions into suitable online questions. When doing so, there are two main differences that can be introduced:

- The effect of the delivery and answering mechanism (i.e. reading from the screen and using a keyboard and mouse rather than a pen and paper).
- The effect of any rewording of the question.

These are not just issues associated with the quality and design of specific questions, but with the wider area of comparability of paper and computer-based assessment, and can include differences due to attitude and behaviour as well as the physical reading and answering mechanisms of the screen, keyboard and mouse.

The work described here is part of a growing body of research into the validity of online assessment in a range of subjects and at various levels of education. This includes research into the effects of medium on multiple-choice questions (i.e. Lee & Weerakoon, 2001), on questions utilising multimedia (for example, see Figure 2), which may be further removed from their paper counterparts (i.e. Greenwood et al, 2000; Sim & Horton, 2005), as well as research into the effects on different subject areas and educational levels: for example, on how the medium effects the performance of primary/elementary school mathematics (Johnson & Green, 2004a), and the impact on the perceptions, behaviour and strategies of these students (Johnson & Green, 2004b). To date, the conclusions of these studies have been varied.

In order to investigate the two issues of medium and rewording in the PASS-IT project, experiments were designed to isolate any potential differences due to these factors. Investigating the medium and rewording effects involved the creation of three versions of the assessment:

- the original paper NAB assessment (called P),
- the computer version of the assessment (called Q), and
- a screen dump of the computer version (called R – Reverse translation).

The computer versions of the assessments were approved by SQA subject specialists prior to their use – this being normal practice for the paper NABs. Quality assurance procedures also included post moderation of the computerised marking by SQA external moderators. Investigation into the medium effect was carried out by comparing Q & R, and into any rewording effect by comparing P & R. Group allocation and the statistical model took ability, gender and teaching class into consideration. This approach had been previously used in earlier work by Fiddes et al (2002).

Delivery Medium

Previous work by Fiddes et al (2002) investigated Mathematics assessments at the SQA level of Higher, and showed that there was no evidence of a difference due to the delivery medium. As part of PASS-IT the effect of the delivery medium was investigated in the unit assessments in other subject areas – those of Higher Chemistry and Higher National Certificate (HNC) in Computing. Performance data was gathered from 137 Higher Chemistry students, and 44 Higher National Computing students, and the statistical analysis of this data was carried out using a general linear model, taking into account factors such as differences in ability, school and gender – more details on the statistical analysis can be found in Ashton et al (2005).

The results of this research also showed no evidence of a difference due to the delivery medium (Ashton et al, 2005). This lays down good practice for the conversion of valid paper assessments into valid computer assessments.

Rewording

In Higher Chemistry and Higher National Certificate (HNC) in Computing the results of research showed no evidence of a difference due to rewording in these assessments (Ashton et al (2005). However, in Higher Mathematics there was evidence of a difference due to rewording (Fiddes et al, 2002), even though the questions were considered to be assessing the same learning objectives. It was thought that this may have been due to the awarding of partial credit in these examinations – that is the awarding of marks to working or answers that are partially correct.

Partial Credit

Partial credit is often awarded in examinations where a student has not obtained the correct answer, but has satisfied some of the objectives. For example, in Mathematics a student could have demonstrated use of the correct strategy, but a small error early on could result in an incorrect final answer. The awarding of partial credit is often not incorporated into online assessments, where marks are usually given on the basis of a final answer. The influence of partial credit in online assessments for Higher Mathematics was borne out by the work of McGuire et al (2002), where comparison of marking with and without partial credit showed a highly significant difference.

This study also investigated a mechanism to support the awarding of partial credit using a feature of the assessment system called optional steps. Figure 1 shows an example of a question with optional steps, where the steps have not yet been revealed. In this case the candidate can choose to use the steps to answer the question, rather than answering only a single keypart. Figure 3 illustrates the effect of clicking the steps button. When the steps button has been pressed, the candidate must answer all the parts displayed in order to gain the maximum marks. The study by McGuire et al (2002) compared the use of optional steps, compulsory steps and no steps and concluded that optional steps could provide a mechanism to implement the partial credit specified by the marking schemes.

However, McGuire et al (2002) were aware that, in providing these mechanisms, in some cases it could be argued that different learning objectives were being assessed. The research in this area in PASS-IT took this work a stage further by investigating deduction of marks when steps are taken (Figure 3 shows the removal of 0.5 marks in part 8.1.1), and by studying the effect of providing simple feedback throughout the exam. The results again showed that the steps approach could provide a mechanism to deal with the issue of partial credit, but highlighted the importance of accounting for 'follow through' errors (Ashton et al, 2006). Follow through errors are those

where an initial mistake results in an incorrect value being 'followed through' and used in the remaining working, but where credit should be obtained for the method.

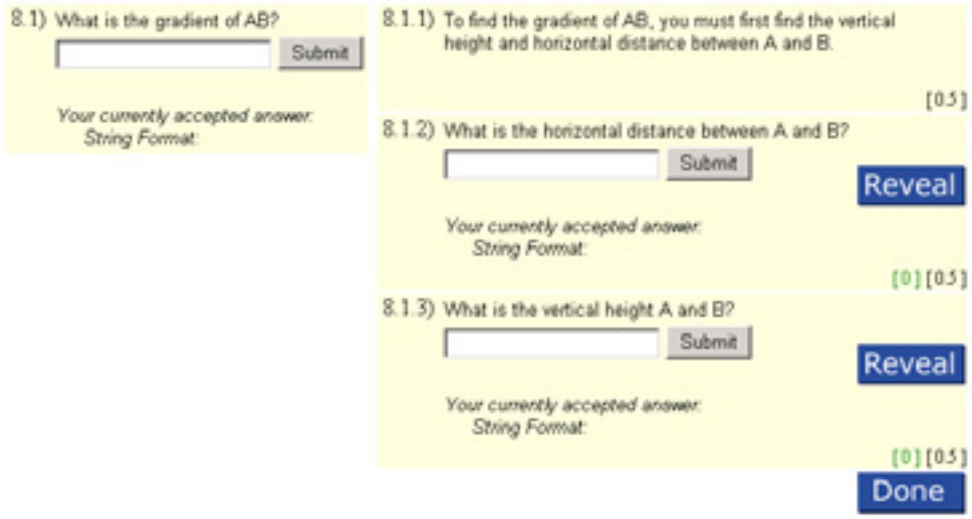

Figure 3: Intermediate 2 Mathematics question with optional steps shown.

Conclusion

This article gives an insight into some of the research and development that were part of PASS-IT. Quality assurance procedures included pre-approval of the assessments and post moderation of the computerised marking by SQA external moderators, and medium and rewording experiments have shown little or no evidence of a difference in attainment, suggesting that these tests were as valid and reliable as the paper-based tests. The work in the project has also shown that in order to have a reliable and valid measure of student performance in Mathematics the assessment system and questions must enable issues such as partial credit to be taken into account.

In addition, the technologies must support the educational requirements of the specific subject and level. Whilst many useful features for mathematical and scientific subjects already exists in the PASS-IT software (for example, randomisation and mathematical equivalence), it is important to explore other subjects in more detail to ensure technology is developed in accordance with the educational needs of these subjects. To this end PASS-IT included subjects such as Music, French and English, and as a result, developed the inclusion of two-way integration of multimedia, which was key to supporting question development in these subjects. However, further exploration of the needs of specific subjects at different educational levels should be carried out.

PASS-IT research also explored the use of different types of questions to measure the same objectives, the automatic marking of answers in the form of free-text (Mitchell & Aldridge, 2003), as well as mathematical expressions, and developed good practice in question design (Ashton & Youngson, 2004). The results of these experiments have provided confidence in the validity of this approach.

In addition, as technology improves, our demands of online assessment may change. Bennett (1998) suggests that we are only in the early stages of our use of computers in assessment, and in the future we will begin to change the way we use computers in assessment, and the nature of assessment, for example, by increasing use of adaptive technology, new question formats, simulations and virtual reality. Following on from PASS-IT, work in this area is continuing, exploring the use of simulations in assessment, from a technical, practical and educational point of view, to further broaden the knowledge, skills and strategies which can be assessed (Thomas et al, 2004, 2005; Thomas & Ashton, 2005; Ashton & Thomas, 2006).

The online learning system SCHOLAR (http://scholar.hw.ac.uk) is being used throughout Scotland for computer-based learning and online formative assessment. The work in this article has moved on to deliver summative assessment and is currently providing online assessment in parts of

the Mathematics and Computing and IT curriculum to over 20 Scottish schools. The wide range of e-assessments developed are being viewed as prototype e-NABs and the research findings will inform the future development and implementation of online assessment for SQA qualifications.

Acknowledgements

PASS-IT was a collaborative venture with contributions from staff in the five organisations and many staff and students in Scottish schools and colleges. Our thanks go to all involved for their hard work and cooperation.

Correspondence

Dr Helen Ashton, School of Mathematical and Computer Sciences, Colin MacLaurin Building, Heriot-Watt University, Riccarton, Edinburgh EH14 4AS, United Kingdom (h.s.ashton@hw.ac.uk); Christine Wood, CAA, Scottish Qualifications Authority, The Optima Building, 58 Robertson Street, Glasgow G2 8DQ, United Kingdom (christine.wood@sqa.org.uk).

References

Ashton, H. & Thomas, R. (2006) Bridging the Gap Between Assessment, Learning and Teaching, *Proceedings of the 10th International CAA Conference*, Loughborough: Loughborough University. http://www.caaconference.com/pastConferences/2006/proceedings/ashton.pdf

Ashton, H.S. & Youngson, M.A. (2004) Creating Questions for Automatic Assessment in Mathematics, *LTSN MSOR Maths CAA series*, February article. http://ltsn.mathstore.ac.uk/articles/maths-caa-series/feb2004/

Ashton, H.S., Beevers, C.E., Korabinski, A.A. & Youngson, M.A. (2005) Investigating the Medium Effect in School Chemistry and College Computing National Examinations, *British Journal of Educational Technology*, 36(5), pp. 771-787. http://dx.doi.org/10.1111/j.1467-8535.2005.00501.x

Ashton, H.S., Beevers, C.E., Korabinski, A.A. & Youngson, M.A. (2006) Incorporating Partial Credit in Computer-aided Assessment of Mathematics in Secondary Education, *British Journal of Educational Technology*, 37(1), pp. 93-119. http://dx.doi.org/10.1111/j.1467-8535.2005.00512.x

Ashton, H.S., Beevers, C.E., Schofield, D.K. & Youngson, M.A. (2004) Informative Reports – experiences from the PASS-IT Project, *Proceedings of the 8th International CAA Conference*, Loughborough University. http://www.caaconference.com/pastConferences/2004/proceedings/ashton.pdf

Ashton, H.S., Schofield, D.K. & Woodger, S.C. (2003) Piloting Summative Web Assessment in Secondary Education, *Proceedings of the 7th International CAA Conference*, Loughborough University. http://www.caaconference.com/pastConferences/2003/procedings/ashton.pdf

Bennett, R.E. (1998) Reinventing Assessment: Speculations on the Future of Large Scale Educational Testing, *Educational Testing Service*. http://www.ets.org/research/researcher/PIC-REINVENT.html

Fiddes, D., Korabinski, A.A., McGuire, G.R., Youngson, M.A. & McMillan, D. (2002) Are Examination Results Affected by the Mode of Delivery?, *Alt-J*, 10, pp. 60-69.

Greenwood, L., Cole, U.M., McBride, F.V., Morrison, H., Cowan, P. & Lee, M. (2000) Can the Same Results Be Obtained Using Computer-Mediated Tests as for Paper-Based Tests for National Curriculum Assessment? *International Conference on Mathematics/Science Education & Technology*, 1, pp. 179-184.

Johnson, M. & Green, S. (2004a) Online Assessment: the impact of mode on student performance, British Education Research Association Annual Conference, Manchester.

Johnson, M. & Green, S. (2004b) Online Assessment: the impact of mode on students' strategies, perceptions and behaviours, British Education Research Association Annual Conference, Manchester.

Lee, G. & Weerakoon, P. (2001) The Role of Computer-Aided Assessment in Health Professional Education: a comparison of student performance in computer-based and paper-and-pen multiple-choice tests, *Medical Teacher*, 23, pp. 152-157. http://dx.doi.org/10.1080/01421590020031066

McGuire, G.R., Youngson, M.A., Korabinski, A.A. & McMillan, D. (2002) Partial Credit in Mathematics Exams: a comparison of traditional and CAA exams, *Proceedings of the 6th International CAA Conference*, Loughborough University.

Mitchell, T. & Aldridge, N. (2003) Computerised Marking of Free-Text Responses: PASS-IT Project Report for the Scottish Qualifications Authority.

Schofield, D. & Ashton, H. (2005) Effective Reporting for Online Assessment: shedding light on student behaviour, *LTSN MSOR Maths CAA series*, February article. http://mathstore.ac.uk/articles/maths-caa-series/feb2005/

Sim, G. & Horton, M. (2005) Performance and Attitude of Children in Computer Based Versus Paper Based Testing, World Conference on Educational Multimedia, Hypermedia and Telecommunications, Montreal.

Thomas, R. & Ashton, H. (2005) Use and Reuse of Simulations in Mathematical Online Assessment, *LTSN MSOR Maths CAA series*, July article. http://mathstore.ac.uk/articles/maths-caa-series/july2005/

Thomas, R., Ashton, H., Austin, B., Beevers, C., Edwards, D. & Milligan, C. (2004) Assessing Higher Order Skills Using Simulations, *Proceedings of the 8th International CAA Conference*, Loughborough University. http://www.caaconference.com/pastConferences/2004/proceedings/Thomas_R.pdf

Thomas, R., Ashton, H., Austin, B., Beevers, C., Edwards, D. & Milligan, C. (2005) Cost Effective Use of Simulations in Online Assessment, *Proceedings of the 9th International CAA Conference*, Loughborough University. http://www.caaconference.com/pastConferences/2005/proceedings/ThomasR_AshtonH_AustinB_BeezersC_EdwardsD_MilliganC.pdf