PILOTING SUMMATIVE WEB ASSESSMENT IN SECONDARY EDUCATION

Helen S Ashton, David K Schofield and Susan C Woodger

Piloting Summative Web Assessment in Secondary Education

H S Ashton, D K Schofield and S C Woodger

School of Mathematical and Computing Sciences, Heriot-Watt University, Riccarton, Edinburgh EH14 4AS.
Tel: 0131-451-3251, Fax: 0131-451-3249

h.s.ashton@hw.ac.uk; d.k.schofield@hw.ac.uk; s.c.woodger@hw.ac.uk

Abstract

This paper describes a major collaboration between several educational agencies in Scotland to pilot the automatic delivery of a number of national qualifications directly into schools and colleges.

The agencies of the Scottish Qualifications Authority (SQA), the Scottish Further Education Unit (SFEU), Learning and Teaching Scotland (LTS), BBC Scotland and the recently formed Scottish Centre for Research into On-Line Learning and Assessment (SCROLLA) have combined to research the issues surrounding automatic delivery of internal assessments in Chemistry at Higher and Advanced Higher, Computing at Higher National level and Mathematics across the entire curriculum. These internal assessments are designed to test the minimum competencies or lower order skills in the relevant subject.

The process involves collaboration between traditional examiners from the SQA and learning technologists at LTS, SCROLLA and the SQA in an attempt to devise e-questions that measure the learning outcomes of a number of courses. The automatic web assessment system developed at Heriot-Watt University over many years is being employed since it has the range of question types necessary to handle such a diverse group of subjects. In fact, new question types continue to be introduced, for example the technical advance of integrating Flash™ with the automated marking system was accomplished some months ago.

This paper will explain the educational and technological progress of the project known as PASS-IT (PASS-IT), which is funded by the Scottish Executive. This paper will discuss some of the issues that have arisen in the light of a number of pilots in Edinburgh and Aberdeen schools and colleges across Scotland in Spring 2003. If successful, PASS-IT has the potential to provide a blueprint for the roll-out of innovative automatic delivery of summative assessment for the measurement of the lower order skills in a number of academic subjects, building on previous work (Fiddes et al, 2002; McGuire et al, 2002; Beevers and Paterson, 2003).

Introduction

The Scottish educational system undertook some major changes in the year 2000 and introduced the notion of internal testing in a range of national qualifications. Typically, the new Higher and Advanced Higher courses (roughly AS and A Level in England) consist of three units, one per term. In order to be eligible for the end-of-course examination Scottish pupils must sit and pass an internal test at the end of each of the three units of the course. These internal tests are set to examine the minimum competencies in the subject and as such are measuring the lower order skills as defined by, for example, Bloom et al (Bloom et al, 1956,1964). Bloom's taxonomy of learning skills enumerates the levels as

Level 1: Knowledge;
Level 2: Comprehension;
Level 3: Application;
Level 4: Analysis;
Level 5: Synthesis; and
Level 6: Evaluation.

Such taxonomies exist in most subjects with some variations of interpretation, see, for example, the eight point hierarchy advocated by Smith et al (Smith et al, 1996) for Mathematics.

The PASS-IT Project started on 1st August 2002 with a remit to pilot the delivery of some of the National Assessment Bank (NAB) of questions available for internal testing in Scottish schools and colleges. In phase 1 this delivery is limited to the subjects of Chemistry and Mathematics at Higher and Advanced Higher and for some of the units in Higher National Computing. The project recruited three learning technologists to be located at the Scottish Qualifications Authority (SQA), Learning and Teaching Scotland (LTS) and at Heriot-Watt University in Edinburgh. Together with traditional examiners employed by SQA these three computing officers produced electronic questions in a range of subjects for pilots in seven schools and three Further Education colleges. The software used (CUE) is the internally developed web assessment system at Heriot-Watt University that has gone through a series of metamorphoses to emerge with a range of question types well suited to the delivery of questions in many subjects

What are the educational issues?

The challenge for on-line assessment is not a technical one, but a pedagogical one. There are a number of issues, for example: Does the medium matter? Are paper based questions of the same difficulty as on-line questions? Does an on-line assessment measure the same learning outcomes as a traditional paper based assessment? If investigations into assessment are to continue these issues must be addressed. The PASS-IT Project aims to investigate some of these issues through a number of studies.

The analysis for phase 1 has yet to be completed, however, certain issues have come to light from the experiences of authoring and delivering materials for phase 1, and from previous work at Heriot-Watt University. This paper will briefly explain the main aspects of the PASS-IT Project, focusing on the issues of acknowledging partial credit, designing questions that can address the same learning outcomes as a paper based assessment, and how accessibility and special educational needs have been addressed.

Awarding partial credit

In Mathematics the work of PASS-IT has built on earlier research between the CALM (Computer Aided Learning in Mathematics) team at Heriot-Watt University and the SQA (Fiddes et al, 2002; McGuire et al, 2002). In earlier work the effect of the medium and the role of optional steps has been investigated. Particular emphasis on the use of optional steps in the earlier trials appears to offer a way of providing for some partial credit to students of Mathematics. The provision of optional steps in typically multi-stage questions is clearly beneficial in formative assessment to a group of students unable to proceed initially. The ability of the assessment system to generate scaffolding for progression is more desirable than simply giving a hint since it then builds on prior knowledge. However, the use of optional steps to allow progression in summative testing may provide too much information and give away, for example, the strategy in certain questions.

In traditional paper based assessments students have the ability to record their answers and processes on paper. When a human marker marks a question they take into account any rough working made by the student. Similarly, if an error is made in the answer to one part of a question, and that answer is subsequently needed for another part, so long as the correct method has been used the student will not be penalised for the same error twice (follow through). The PASS-IT Project will investigate the use of steps as a method for gaining partial credit in phase 1 and has specifically included steps in questions to facilitate this. An investigation of follow through is planned for phase 2.

The ability to include optional steps enables a question to be posed in a variety of ways. Consider the example:

> A greengrocer buys a bag of nine oranges for £ 1.50 and sells them on at 29 pence each. What is the percentage profit on the sale of all nine oranges?
>
> 1. What is the percentage profit? 74

Figure 1: Traditional question

This version of the question is fine for a good student who can make the intermediate calculations needed to complete the answer and gain the full two marks on offer but it does not help the average student during formative assessment exercises. Nor does it provide any opportunity for partial credit in a summative test. A better version of the question can be created with optional steps and would appear as follows after pressing the steps button:

A greengrocer buys a bag of nine oranges for £ 1.50 and sells them on at 29 pence each. What is the percentage profit on the sale of all nine oranges?

1.1 How much in pence is there when all nine oranges are sold? 261
1.2 What is the profit in pence when all nine oranges have been sold? 111

1. What is the percentage profit? 74

Figure 2: Optional steps

In this form of the question steps 1.1 and 1.2 carry 0.5 marks each and the key part itself 1 further mark so that again two marks are possible. This time, however, the student has to calculate and submit three answers in order to gain full marks. The benefit in formative assessment is that the student can progress when stuck on how to make the first step. Moreover, with ticks and crosses visible on screen the student can gain confidence as the answers are entered to steps 1.1 and 1.2. In summative testing the device of optional steps provides for some partial credit, even if the student cannot make the final step of calculating the percentage profit. One other variant of this approach would be to provide the student who is stuck with a hint called an information step, which does not seek an answer. This could appear alone, or alongside steps requiring an answer. For example, on pressing the optional steps button the question may appear as:

A greengrocer buys a bag of nine oranges for £ 1.50 and sells them on at 29 pence each. What is the percentage profit on the sale of all nine oranges?

1.1 First calculate how much in pence there is when all nine oranges are sold.
1.2 What is the profit in pence when all nine oranges have been sold? 111

1. What is the percentage profit? 74

Figure 3: Question with information step

This time no marks are given for the information step 1.1 and one mark for step 1.2 with a further mark for the key part 1 itself giving again two marks for completing this question. This is again good practice for formative assessment where the emphasis is to provide feedback and encourage the student to make progress.

It could be argued though that this does not properly reward the good student in a summative exercise. Even though the average student has taken longer to complete this question through the steps, and may therefore have put some additional pressure on himself/herself to complete the whole timed test, nevertheless, the more traditional examiner might wish to differentiate between candidates more subtly. So, finally, consider the example in Figure 3 again.

This time a more traditional examiner might argue that information step 1.1 gives away the strategy for this question so that 0.5 marks should be deducted for this hint. However if step 1.2 is still worth one mark and key part 1 an additional mark then the student can score 2 - 0.5 = 1.5 marks for a correct solution. A candidate achieving the correct answer without steps would still gain the full 2 marks. This approach helps to differentiate between candidates and adds to the sophistication of automatic testing.

Are we assessing the same learning objectives?

Another issue is exemplified by a question from the NAB for Higher Mathematics in which the student is asked to show the effect of a specific operation on a function. The traditional way of exhibiting appropriate understanding is to draw the transformed function given a diagram of the initial function (Figure 4).

Diagrams 1 and 2 on the worksheet show part of the graph y=f(x).
(a) On Diagram 1, draw the graph of y=-f(x).
(b) On Diagram 2, draw the graph of y=f(x+3).

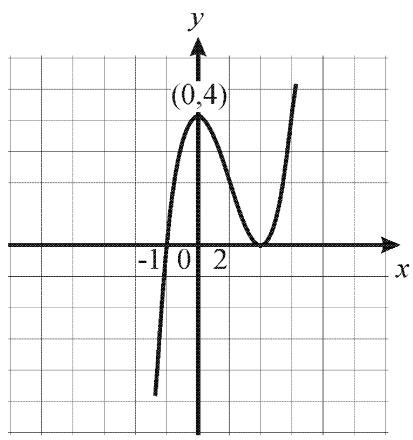

Diagram 1

Figure 4: Original paper based question.

A traditional examiner would be looking for the student to have sketched the original graph reflected in the x-axis. One option for an on-line version would be to use multiple choice, presenting the student with a number of options, possibly combinations of the reflection in each axis. However, this reduces the difficulty of the question from one of the application of knowledge, to recognition and elimination, deemed wholly unsatisfactory and not equivalent to the traditional way of measuring this understanding. In this case the solution developed for the PASS-IT phase 1 pilots was to use a hotspot question, where each integer coordinate on the graph was a hotspot. The task then became one of clicking on all the turning points and the points of intersection of the x-axis. A selected hotspot was shown as a red circle, and marks were only awarded for all points being correct. Figure 5 shows the students' view after a correct answer has been given.

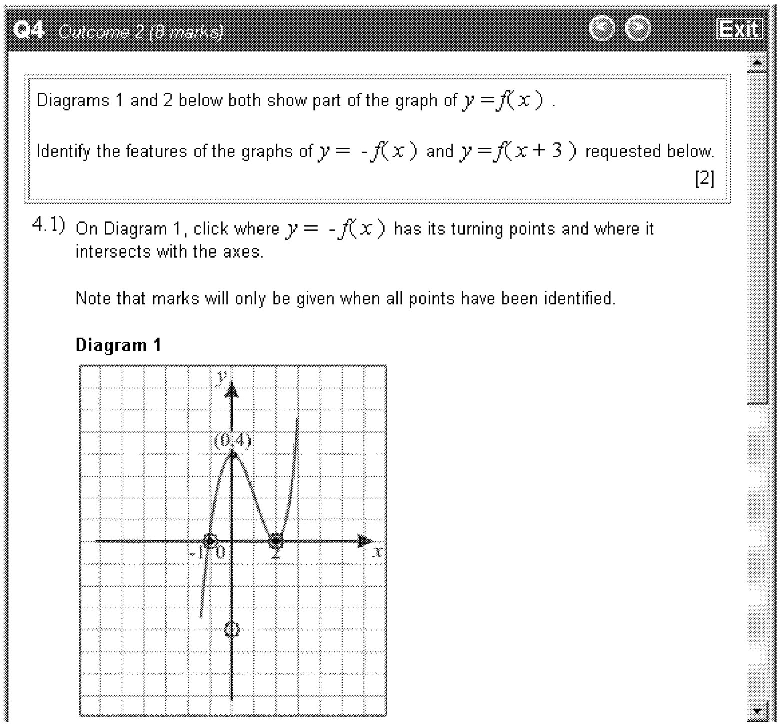

Figure 5: A correct answer to a Hotspot version of the question

This solution removes the problem of only testing recognition as it allows any combination of points to be chosen.

However, ingenious though this solution is, it still poses some problems:

- The student never sees/draws the new function.
- Clicking on the exact point can be difficult, and may reduce accessibility.
- The question takes a long time to produce (this one has 169 hotspots!).
- The question has a large file size, which could pose problems with server load.
- Changes could be tedious, and there is no opportunity for randomisation.

From the students' perspective, not seeing the final function could be a huge barrier to identification of the correct answer, and to learning from that. This problem is likely to be more pronounced in lower school years. In addition, the hotspots themselves must be small so that unique identification can be made, and to keep the image size to an acceptable level. Students with difficulties using the mouse, or where a mouse is "sticky" (a common problem), could have problems selecting the correct hotspot. Every choice the student makes is communicated to the server, allowing the corresponding highlighted circle to be shown. This takes time, even with the speediest of connections. Combine this with time to deselect incorrectly chosen options, and this question could take a considerable amount of time.

From a technical perspective, a question with 169 hotspots has to contain a reasonable amount of information, increasing the file size quite substantially. As a comparison, this file was ~60KB (not including the image). The next largest file in this NAB unit was only 15KB. Although 60KB is not huge, consider this for even a single class of students, with multiple hotspots to select, and quickly potential traffic and server load problems may arise.

From an authors' perspective the question takes a significant amount of time to produce, and can be tedious to change. There is also no potential to randomise the function being presented (something that is useful in formative assessment, and may be useful to combat the close proximity of computers in classrooms).

One of the recent technical developments in the assessment engine used in the PASS-IT project was to implement two-way communication with multimedia elements such as Macromedia Flash™. This enables a new approach to this question (see Figure 6).

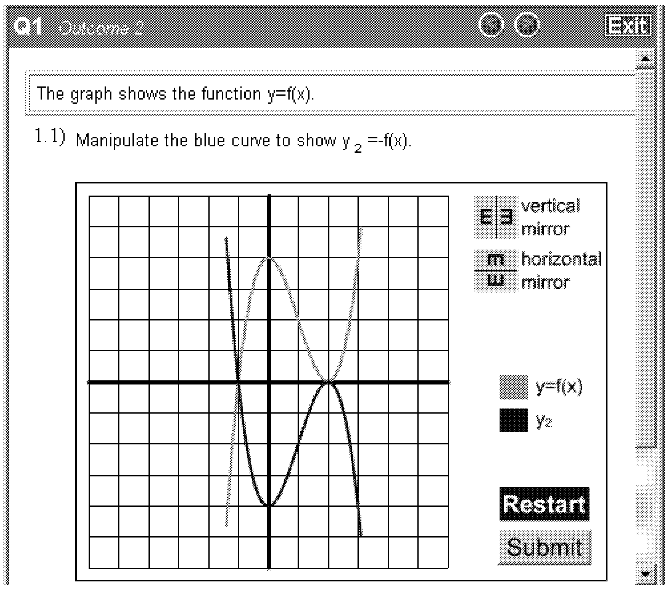

Figure 6: A multimedia version of the question

This solution addresses all of the issues raised above. This solution allows the student to perform translation and reflection operations on the original curve. As with the hotspot solution, the student can position the new function in many locations. The original function is always visible (the lighter of the two curves), however, the student can now also see the new function, improving familiarity and the potential for evaluating their answer and learning from the assessment.

Translations can be performed by using the mouse to drag the curve. The "live" area to drag extends well beyond the original curve, and when "dropped" the curve will snap to the closest grid point, reducing mouse related problems. Reflection operations can be performed using two buttons (top right). In addition, keys can be used for translation (arrow keys) and mirror operations (h and v), further improving accessibility.

From the technical perspective, the combined file size of the Flash™ element and the question is now ~10KB. The answer is only communicated to the server when the student chooses to submit their answer, further reducing traffic and server load.

From the authors' perspective, once the Flash™ element has been produced, the authoring and modification of questions is dramatically improved. Communication between the assessment engine and Flash™ allows many parameters to be sent in, allowing both extensive reuse, and randomisation of the question.

There are many other examples where appropriate use of integrated technology could enhance the range of potential objectives that could be measured, the flexibility of generation and reuse, and the usability and accessibility of the system. It is hoped that phase 2 of the project will allow further exploration of the use of multimedia.

Accessibility and Assessment

For widespread acceptance of an on-line assessment system it needs to be at least as accessible as current paper based systems. In addition to pedagogical issues the PASS-IT Project is also investigating accessibility issues with on-line assessment.

System Accessibility

The assessment system has been designed to be accessible. The W3C Web Accessibility Initiative (WAI) Web Content Accessibility Guidelines (W3c a, b, c), were widely consulted as was the TechDis Web Accessibility and Usability Resource (TechDis), a JISC service which provides information on the practical aspects of accessibility and its evaluation. Many of the system's features will enable users with dyslexia, vision impairment and reduced mobility to use the system.

The assessment system embraces the Web Accessibility Guidelines, for example, through the use of relative sizing for fonts and tables and individual stylesheets to give greater control to the user over the presentation of the assessment.

The screen-reading package JAWS is able to read out the questions to a vision impaired student. All mathematical expressions are rendered with a string equivalent. The string equivalent is placed in the *alt* tag enabling non-visual browsers to access the expression. Naturally, the system also makes use of other tags such as *title* and *abbr* to present information in accessible forms.

Other features increase accessibility for a wide variety of users. For instance, shortcut keys are employed to allow navigation within the assessment system without the use of a mouse. The use of multimedia elements, for example Flash™, also offers opportunities to facilitate questions in a format that supports accessibility. With the graph question described above, a vision impaired student could have the key coordinates of the graph read out.

Combine this with the ability to use keys to position and manipulate the graph and the accessibility of the question is dramatically improved.

PASS-IT Project On-line Assessments

As the pilots have progressed care has been taken to address any accessibility issues that have arisen. Students who required additional help, due to special educational needs, were identified. The majority of educational needs were due to dyslexia and affected male students only.

This is in line with the statistic that dyslexia is 4 times more common in males as it is in females and affects 4% of the population severely, as published by The British Dyslexia Association (BDA). In the same article The British Dyslexia Association states that "1 child in every classroom will need ongoing appropriate specialist teaching throughout his/her time in school and support in further education, training and employment".

An example of the types of problems encountered by dyslexic students is with letters moving about on the printed page and on screen. Improvements are possible by reducing the glare on paper and computer screens and using different colour combinations. In discussions, students were generally unaware that changing the background colour of web pages was possible. Only one student was found to be aware of stylesheets and changed his textboxes to a green background to relieve the exhaustion of looking at a white screen. Another student found that wearing green tinted spectacles worked well. As the SQA provides green paper for dyslexic students, the Project developed a stylesheet allowing students to override default settings to turn the screen background green.

One student required a scribe during paper tests due to problems with both reading and writing. Using the assessment system the student was able to employ exactly the same routine as he was used to in paper tests. He was keen to experience using the on-line assessments, as, with practice, he felt it had the potential to eliminate the difficulties he encountered with handwritten tests. In fact, in their feedback, many students commented upon preferring typing as it meant they did not have to worry about the legibility of their handwriting.

What are the next steps?

The next phase of the project will present some new challenges, as, in addition to other subjects, assessments will be carried out with children in the 5-14 age range. The use of partial credit would be greatly enhanced with the implementation of follow through (Ashton and Beevers, 2002). It is anticipated that this will be of greater importance in the lower school years, as will an increased use of multimedia components to address specific learning outcomes, and to enhance engagement. There is also the potential for exploring the use of different presentation styles with younger children to improve usability and engagement.

Planning for phase 2 of the PASS-IT project is currently under way and it is hoped that there will be the opportunity to explore some of these issues further.

Acknowledgements

The authors wish to acknowledge the contributions of all members of the PASS-IT project team, and more specifically to thank Professor Cliff Beevers for his comments during the preparation of this article.

References

Ashton, H. Beevers, C.E. (2002) *Extending Flexibility in an Existing On-line Assessment System* Sixth International CAA Conference, Loughborough.

BDA *Dyslexia research information* <http://www.bda-dyslexia.org.uk/main/information/extras/x02stats.asp>

Beevers, C.E. McGuire, G.R. Stirling, G. and Wild, D.G. (1995) *Mathematical Ability Assessed by Computer* J Comp. Educ. 25, 123 - 132.

Beevers, C.E. Youngson, M.A. McGuire, G.R. Wild, D.G. and Fiddes, D.J. (1999) *Issues of Partial Credit in Mathematical Assessment by Computer* Alt-J 7, 26 - 32.

Beevers, C.E. and Paterson, J.S. (2003) *Automatic Assessment of Problem Solving Skills in Mathematics* Active Learning in Higher Education (to appear).

Bloom, B.S. Englehart, M. Furst, E. Hill, W. Krathwohl, D. (1956) *Taxonomy of Educational Objectives: the classification of educational goals Handbook 1* London: Longman.

CUE *CUE Assessment System* <http://www.calm.hw.ac.uk/cue.html>

Fiddes, D.J. Korabinski, A.A. McGuire, G.R. Youngson, M.A. and McMillan D. (2002) *Does the mode of delivery affect Mathematics examination results?* Alt-J 10, 60-69.

Krathwohl, D. Bloom, B.S. Masia, B.B. (1964) *Taxonomy of Educational Objectives: the classification of educational goals Handbook 2* London: Longman.

McGuire, G.R. Youngson, M.A. Korabinski, A.A. and MacMillan, D. (2002) *Partial Credit in Mathematics Exams – a Comparison of Traditional and CAA Exams* Sixth International CAA Conference, Loughborough.

PASS-IT Project <http://www.pass-it.org.uk/>

Smith, G. Wood, L. Coupland, M. and Stephenson, B. (1996) *Constructing mathematical examinations to assess a range of knowledge and skills* International Journal of Mathematics Education in Science and Technology, **27**(1), 65-77.

TechDis, *Web Usability and Usability Resource*
<http://www.techdis.ac.uk/seven/>

W3c (a), *W3c Web Content Accessibility Guidelines 1.0*
<http://www.w3.org/TR/WAI-WEBCONTENT>

W3c (b), *W3c Web Content Accessibility Guidelines 2.0*
<http://www.w3.org/TR/WCAG20>

W3c (c), *W3c Authoring Tool Accessibility Guidelines 2.*0
<http://www.w3.org/TR/ATAG20>