Computer-aided assessment in mathematics: Panacea or propaganda?

Duncan Lawson
School of Mathematical and Information Sciences, Coventry University, Coventry CV1 5FB, United Kingdom

Abstract

The proponents of computer-aided assessment are very persuasive in extolling the benefits and virtues of this technology. Undoubtedly, it has many positive features. However, it also has some limitations and drawbacks which need to be considered. Mathematics as a discipline has certain specific problems which are not relevant to the use of computer-aided assessment in many other disciplines. These problems often feature the introduction of the assessment of other learning outcomes into the process. This can create problems in summative assessment but is less of a problem in formative assessment. Evidence is supplied that students find the use of computer-aided assessment in a formative manner to be worthwhile. Finally, a key question about the future development of computer-aided assessment in mathematics is posed.

Introduction

Computer-aided assessment has a number of indisputable benefits which include:

- availability - students can take assessments whenever they can access the network on which the tests are mounted;
- immediate feedback - as soon as students have completed their test they are given, not only their mark, but also feedback (as detailed as the test author wishes) on the questions they answered incorrectly;
- repeated practice - provided a question bank of similar questions has been authored, students can take tests on the same topic repeatedly to develop their skills; and
- anonymity - some students feel more secure giving answers to an inanimate machine than to a human being - there is less embarrassment in giving a foolish answer when it is only the machine that 'knows'.

The proponents of computer-aided assessment will extol these undoubted virtues. However, there are some pitfalls with computer-aided assessment, particularly in mathematics, which academics need to consider before they decide whether it is worth investing their time in preparing and using computer-aided assessment with their students.

In the United Kingdom, the Quality Assurance Agency, the body responsible for the quality of higher education programs, has steered universities firmly in the direction of constructive alignment with a greater emphasis than ever before on the role of assessment[1]. Programs are required to demonstrate that all the intended learning outcomes for the program are assessed. Indeed, failure to assess one of the learning outcomes is one of the few specifically identified grounds on which reviewers would deem the standards to be inadequate.

The QAA approach implies a hierarchical structure. Program outcomes are secured by a combination of the outcomes of the modules or units which make up the program. The assessment scheme for each module must ensure that all the intended outcomes of the module are assessed. This requires that each assessment task within the module should identify the outcomes it is assessing, to ensure that all outcomes are assessed. Of course, some module outcomes may be assessed by more than one assessment task and some program outcomes may be assessed in more than one module.

The use of computer-aided assessment as a delivery mechanism for a specific assessment task should ideally have no effect on the outcome being assessed. However, this does not automatically happen and care must be taken to ensure that using computer-aided assessment does not introduce extra outcomes to the assessment task or allow the assessment to be completed satisfactorily in a way which does not require the achievement of the outcome being assessed.

In subsequent sections examples where the use of computer-aided assessment may cause problems are highlighted. In most cases these problems are not insurmountable, but it is vital for the sake of the reliability of the assessment that they are considered.

Multiple choice and guessing

The easiest question type to implement in computer-aided assessment is multiple choice. Multiple choice questions do not require a complex software package for their implementation - indeed they can be constructed quite easily using standard HTML forms.

Some academic staff have reservations about the use of multiple choice questions. They argue that candidates who have no understanding at all of the subject matter of the question have a chance of getting the question right by simply selecting at random one of the choices offered. Although, in principle, increasing the number of options available reduces the likelihood of a successful guess, the difficulty of producing plausible incorrect answers means that in most cases only a small number of options are given. Furthermore, the form of the choices may give candidates, who otherwise did not know how to begin, an indication of the method required to solve the problem.

One method that is often used to combat the possibility of students guessing the correct answer is negative marking. If a multiple choice question has four options (one correct and three incorrect) then one mark is awarded for a correct answer, whilst one-third of a mark is deducted for an incorrect answer. The idea behind this is that the expected value of a student answering the question at random is zero. This is attractive, however it does penalise the student who genuinely makes a mistake. In a pen and paper examination a student who makes a mistake and so produces an incorrect answer would, at worst (there may be some method marks), be given a mark of zero. In a multiple choice examination with negative marking this student is actively penalised for her/his mistake.

Proponents of multiple choice testing counter the charge that guessing may invalidate the assessment by pointing out the importance of having a reasonably large number of test items. The following is a quote from Blueprint for CAA[2]:

'It's worth remembering that the relevance of guessing decreases as the number of test items increases. If you have a true/false test containing ONE question, a student has a 50% chance of getting it right and scoring full marks. The chances of scoring 100% on a 45 question true/false test are less than one in a trillion (10^{12}) and the chances of earning at least 70% in a 45-question test are less than one in 300.'

In actual fact these are the probabilities for a 40 question test (for a 45 question test the chances are respectively less than one in 35 trillion and less than one in 800).

However, such arguments obscure the point. Most students do not randomly guess all questions but only some. Instead of considering a student who knows nothing and guesses everything, it may be more appropriate to consider the fate of marginal students (i.e. those who would just pass, or just fail, a pen and paper test). In particular, it is more revealing to examine the likelihood of marginal students passing a multiple choice test when they would have failed a pen and paper test, rather than the chances of a complete random guesser achieving 70% or 100%.

Consider two weak students taking a 40 question test with a 40% pass mark. Student A has only focused on a small part of the syllabus and so only knows how to attempt 17 of the questions of which 14 are answered correctly. Student B has focused on a slightly larger part of the syllabus, attempting 24 questions, although only 16 of these are answered correctly.

If this is a standard pen and paper test with one mark for a correct answer and zero otherwise (no fractional marks for method) then student A scores 35% and fails whilst student B scores 40% and passes.

Now look at the fate of these students taking the same questions but taken as multiple choice questions. We will assume that the questions they answered incorrectly are still answered incorrectly (i.e. the presence of the multiple choice options does not help them to correct some of their incorrect answers) and that they randomly guess the answers to the questions they do not know how to attempt.

If the straightforward zero/one marking scheme is retained student A needs to guess correctly the answers to two questions out of 23 in order to achieve a pass mark of 40%. With four multiple choice options the probability of doing this is almost 99%.

If negative marking is employed, so that the student is penalised by one-third of a mark for each wrong answer, the student needs to gain three marks from guessing in order to pass. This requires at least eight correct answers from the 23 guesses. Clearly this is a less likely event; but it still occurs 20% of the time.

Student B, when taking the test with pen and paper was the most marginal pass student possible. When taking this test using zero/one multiple choice the student has a 60% chance of gaining a

mark of over 50% (by correctly guessing the answers to at least four out of 16 questions). However, if negative marking is used, this student (who is clearly better prepared than student A) is in considerable difficulty. The eight incorrect answers and consequent loss of 8/3 marks means the student must guess correctly the answers to six of the remaining 16 questions in order to regain the lost marks and register a pass. This will happen 19% of the time. So 81% of the time student B will fail the test although 40% of the questions were answered correctly from the student's knowledge and not because of guessing.

The fates of student A and student B demonstrate some of the inherent problems associated with multiple choice tests.

Introduction of extra learning outcomes

The software available to deliver computer-based assessment has advanced significantly over recent years. However, it still has and almost certainly always will have some limitations. Academic staff are creative in finding ways round these limitations - however these 'fixes' can introduce extra learning outcomes into the assessment process.

For example, a standard question to assess the learning outcome 'is able to solve quadratic equations' would be:

Solve the quadratic equation $2x^2 + 3x - 2 = 0$

However, in the computer-aided assessment package produced by Ward and Lawson[3], this question becomes:

What is the larger root of the quadratic equation $2x^2 + 3x - 2 = 0$?

The reason for this change is that *Question Mark Designer*, the software used to produce this assessment package, does not allow questions whose answers are unordered lists. Although the change to the question is only minor it has introduced the implicit assessment of an extra learning outcome into the question. In order to be able to answer the question correctly the student must not only be able to find the two solutions of the equation but also be able to order them correctly. Now it may be argued that this is a relatively simple task that should be well within the compass of anyone who can correctly solve the quadratic equation. However, there is scope for error, particularly when both solutions are negative. Occasionally a student who would have correctly answered the question on paper will give the wrong answer.

There are other simple variations of this problem. Hawkes[4] suggested that one way of delivering a question which required the student to determine the coordinates of the maximum point on a given curve was to ask for the sum of the squares of the coordinates. Again the thinking is that any student capable of using calculus to find the location of the stationary points and then to

determine which is the maximum would be perfectly capable of squaring the two coordinates and then adding the two squares. However, even able students do sometimes make arithmetic slips. Once more a learning outcome has been added in order to allow the question to be delivered by computer rather than on paper.

Both these examples are caused by limitations in the software used to implement them. In mathematically focused software, such as that described by Beevers and Paterson[5], answers which take the form of unordered lists (required for the quadratic equation question) and ordered lists (required for the maximum question) are allowed. With such software these particular problems are solved. However, this does not mean that the problem of the introduction of extra learning outcomes into the assessment process has been eliminated. It simply means that it has been pushed further along (as we shall examine in subsequent sections).

The input of mathematics

Computer-aided assessment in mathematics has two major difficulties that are not present in many other disciplines. The first is how to enter mathematical expressions, particularly when the standard formatting of these expressions requires more than a single line (such as with fractions, powers, etc.). The second is the way in which mathematical expressions are checked to determine if they are correct. As this second difficulty does not impact particularly on the assessment experience it is not discussed further here.

In the early days of computer-aided assessment in mathematics, questions requiring mathematical input used linear input. Such linear input was based broadly on the kind of syntax students would have encountered in spreadsheets[6]. This meant that students had to be competent in this form of representation of mathematical expressions in order to be able to complete successfully computer-aided examinations. Now it is desirable to expect students to use standard mathematical notation and formatting when producing written answers - indeed this is an essential skill that students must have and one that has been implicitly assumed to be an assessed learning outcome of every written mathematics assessment. However, it is completely different to require mastery of non-standard syntax (it is non-standard because different assessment packages used different versions - for example, should exponentiation be denoted by ^ or by **?).

For this reason specifically mathematical assessment packages have developed their own input tools (see, for example, Beevers and Paterson[5]). These tools vary. Some still use linear input but then present the formatted representation of the linear input so that students can check that the input they have given does represent the answer they wish to enter. If it does not then they can edit the linear input and check the new formatted representation and repeat the process until they have successfully entered their answer. Other input tools use palettes and templates in a manner similar to *Equation Editor* within *MS Word*. With this approach the student builds up the formatted version of their answer directly rather than indirectly via linear input. Whilst both of these methods are to be preferred to linear input on its own, they still place extra learning requirements on the student. These learning requirements are nothing to do with the learning outcomes of the program or the piece of assessment. They are simply to do with the mechanism being used to deliver the assessment.

As the entire population becomes ever more computer literate it may well be that this problem will disappear. We do not say that because a pen and paper assessment requires students to be able to write that we have introduced extra learning requirements. So, when everyone is completely familiar with building up formatted expressions from palettes and templates, we will be able to say the same about input tools.

Multiple choice: Another problem

One way of avoiding the need to have an input tool for questions with answers that require formatted mathematical input is to use multiple choice questions. Ward and Lawson[3] use this device in some integration questions.

So, instead of asking the question:

Use the substitution $u = 1 + 3x$ to determine $\int \cos(1 + 3x)\,dx$

the question is presented as:

Use the substitution $u = 1 + 3x$ to determine $\int \cos(1 + 3x)\,dx$

Answer by selecting the appropriate option.

A $\frac{1}{3}\sin(1 + 3x) + c$

B $\sin\left(1 + \frac{3x^2}{2}\right) + c$

C $\frac{1}{3}\cos(1 + 3x) + c$

D $\frac{1}{6}\sin(1 + 3x)^2 + c$

In this question, there is no need for students to enter a mathematical expression. All that is required is for them to determine the integral and then find their answer amongst the options offered. This question may have been designed to assess the learning outcome 'is able to determine indefinite integrals of standard functions of linear functions'.

Even if a student answers a large number of questions like this correctly, the assessor has no guarantee that the student has achieved this learning outcome. The reason for this is that the student may not be determining any integrals. Instead students may simply take each of the choices offered and differentiate them until one whose derivative matches the integrand in the question is found. Such students have learnt some mathematics (namely differential calculus), but not the mathematics supposedly being assessed (namely integral calculus).

Method marks and partial credit

One particularly difficult issue in attempting to replicate pen and paper assessment in mathematics using computer-aided assessment is the allocation of method marks and partial credit. Although an industrial speaker at a mathematics assessment conference[7] once said,

'There are no method marks in industry',
this message has fallen on deaf ears. It is generally accepted throughout the mathematics community that an incorrect answer can still demonstrate the achievement of some learning outcomes and should therefore be rewarded with some (although not all) the marks available.

Method marks are most commonly awarded in questions requiring a multistage solution process. For a simple question such as, 'What is the value of -8 + 3?', method marks would not usually be available. But in a question such as, 'Find and classify the stationary point of y = (x+1)exp(-2x)', method marks and partial credit would usually be available. Method marks would be awarded to a student who demonstrated that they knew (at least some of) the steps to undertake to solve this problem (even if none of them were carried out correctly). Partial credit would be awarded to students who knew the way to proceed and carried out some of the steps correctly.

In order to award method marks and partial credit it is necessary to have some knowledge of how the student has approached the problem and what some of their intermediate results were. One way to do this using computer-aided assessment is to replace the single large problem with a series of smaller questions.

Beevers and Patterson[5] give an example of this approach for a typical question from the school-university interface, namely:

$$\text{Obtain partial fractions for } \quad \frac{x}{x^2-1}, \quad x > 1$$

Students can elect to answer this in stages rather than giving the complete answer initially. In this approach they are prompted with a series of steps, each of which they must answer in order to determine the solution of the original problem. In this example the steps are:

STEP 1: Factorise the denominator of $\frac{x}{x^2-1}$

STEP 2: If $\frac{x}{x^2-1} = \frac{a}{x+c} + \frac{b}{x+d}$ where c < d, then by finding the common denominator, state the right hand side of the equation which equates the numerators

STEP 3: By equating coefficients of x, state the expression equal to 1

STEP 4: By equating constants, state the expression equal to 0

STEP 5: What is the value of *a*?

STEP 6: What is the value of *b*?

It is undeniable that this approach does indeed give students the opportunity to achieve partial credit. However, it is also true that this approach has altered the assessed outcomes when compared with this question delivered during a pen and paper examination. Firstly, candidates no longer have to know what partial fractions are as the question now displays the partial fraction decomposition. Secondly, as discussed in the section titled 'Introduction of extra learning ourcomes', an extra learning outcome, that of correctly ordering numbers has been introduced. Thirdly, candidates do not have to know an algorithm for determining partial fractions - instead they simply have to respond to the prompts which steer them through steps which will lead to the determination of the partial fractions. Fourthly, candidates must use the expected method - a student who has learnt to find the coefficients in partial fractions by using the 'cover-up' rule may not know how to respond to some of the intermediate steps that have been built into the delivery of this question. As assessors we may be willing to accept the changes in the outcomes being assessed that the move from pen and paper assessment to computer-aided assessment has made (after all, pen and paper assessment is not perfect); however we should not be ignorant of these changes.

Assessors, irrespective of the medium they are using to deliver their assessment, should be clear about precisely what they are seeking to assess. Consider again the question, 'Find and classify the stationary point of $y = (x+1)exp(-2x)$'. If this is set in a pen and paper examination precisely what is being assessed? To answer the question correctly a candidate must find a derivative using the product rule, solve an equation by factorisation, find another derivative by using the product rule, evaluate an expression and use information about the second derivative to determine the nature of the stationary point. Furthermore they must know that these are the steps to go through in order to solve this problem. It is this ability to know and work through a set of connected steps that is at the heart of this assessment. All of the skills required for the individual steps can be assessed with other questions. For example, if all that was to be assessed was that a student knew how to determine the nature of a stationary point then it would be better to ask the question, 'If $dy/dx=0$ when $x=a$, how can you determine the nature of the stationary point at $x=a$'. This has implications for the computer-aided assessor in terms of how the question is constructed but also for the pen and paper assessor in terms of the allocation of partial credit. Many marking schemes for such a question would award marks for correctly determining each derivative (although this is assessment of the same skill) and also for correct evaluation of the second derivative at the stationary point (which is really a much lower level learning outcome than this question is aimed at). Perhaps there is some merit after all in the view, '*There are no method marks in industry*'!

Assessing higher level outcomes

McCabe et al.[8] have attempted to show that computer-aided assessment can be used to assess some higher level outcomes. They have developed questions aimed at assessing students' abilities to formulate mathematical proofs. One such question gives eight statements that constitute a proof by induction of a property of trees. The statements are given in a random order and candidates are required to put these statements into the correct order. Another proof question is shown in Figure 1.
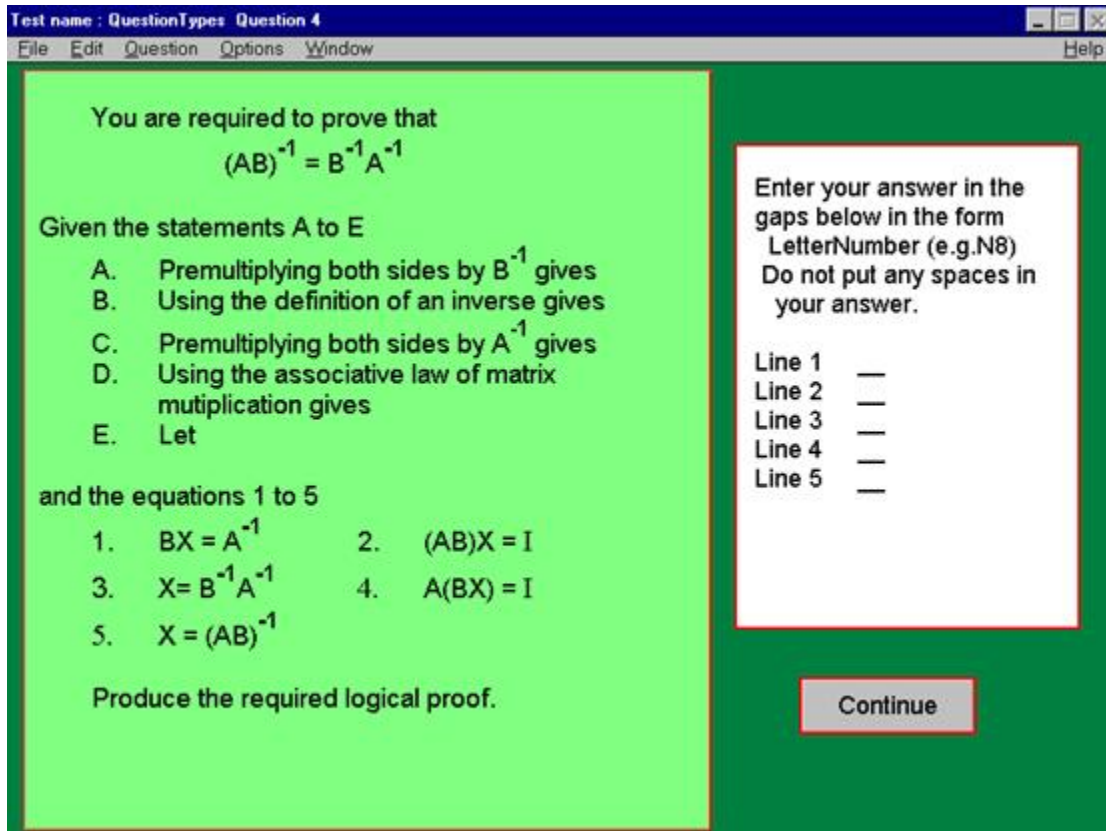
You are required to prove that
$$(AB)^{-1} = B^{-1}A^{-1}$$

Given the statements A to E

A. Premultiplying both sides by B^{-1} gives
B. Using the definition of an inverse gives
C. Premultiplying both sides by A^{-1} gives
D. Using the associative law of matrix mutiplication gives
E. Let

and the equations 1 to 5

1. $BX = A^{-1}$
2. $(AB)X = I$
3. $X = B^{-1}A^{-1}$
4. $A(BX) = I$
5. $X = (AB)^{-1}$

Produce the required logical proof.

Enter your answer in the gaps below in the form LetterNumber (e.g.N8) Do not put any spaces in your answer.

Line 1 __
Line 2 __
Line 3 __
Line 4 __
Line 5 __

Continue

Figure 1. Proof question from McCabe et al.[8]

Whilst this is undoubtedly imaginative use of current technology it certainly cannot be thought of as equivalent to asking a student to prove from scratch that $(AB)^{-1} = B^{-1}A^{-1}$. To assess whether a civil engineer can design a bridge we do not give the engineer a set of components from a dismantled bridge and ask them to indicate how they should be assembled. This is not to say that such questions are without value. Indeed they may serve a very useful purpose in learning some general points about how to construct a proof.

Students who are concerned to maximise their marks rather than their learning may develop strategies for answering questions such as that shown in Figure 1. Such strategies may include things like, 'Proofs often begin with "Let" and then make use of definitions', and, 'When we are trying to find something composite we begin by replacing it with something simpler'. These strategies would help them to identify some of the early stages of this proof as **E5** ('Let' followed by 'X=(AB)$^{-1}$' - the composite that we are trying to prove something about replaced by something simpler), then **Bn** (using the definition). A further strategy could be, 'The proof ends with the

result you have been asked to prove'. This would lead to the conclusion that Line 5 must be **z3**. So far no knowledge of the topic area under consideration has been used but four out of the 10 fragments have been correctly placed. In terms of assessing candidates' ability to prove results about matrix inverses this is not particularly satisfactory. On the other hand, in terms of students learning about the general ideas of proofs, these strategies are very valuable.

Why bother?

Given the list of difficulties associated with computer-aided assessment that has just been outlined, it is reasonable to ask the question, 'Does this mean that computer-aided assessment in mathematics should be abandoned?'. The answer to this question is a definite 'No'. Students derive great benefit from attempting questions and getting immediate feedback on their answers (not just right or wrong but also an indication of how to obtain the correct answer if they were wrong). Practically the only way that this can be delivered on a large scale in higher education today is through the use of computer-aided assessment.

At Coventry University there is a drop-in Mathematics Support Centre which is open for 33 hours per week. Students come to the Centre to receive one-to-one help with problems in mathematics. The Centre also has a web site[9] where students can download copies of all the handouts that are available in the Centre and take practice tests in a range of topics from arithmetic to calculus. Figure 2 shows the usage of the web site during the academic year 2000/01. By far the most used section of the web site is the online test section. In each of the last two months of the academic year almost 800 online tests were taken each month (compared to around 400 in person visits per month to the Centre). As all these test attempts are purely voluntary and completely formative (no record of who attempts the tests or what marks they gain in the test are kept), this indicates that the students certainly perceived value in this computer-aided assessment.
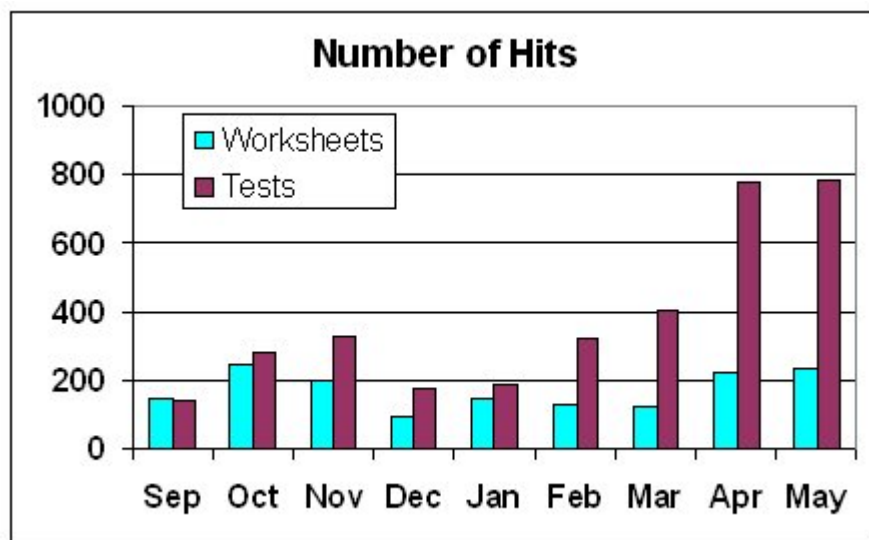

Figure 2. Usage of online tests at Coventry University Mathematics Support Centre web site (from Lawson et al.[9])

Conclusions

The truth about computer-aided assessment in mathematics is that it is neither a total panacea or simply propaganda. It does have limitations but there are circumstances in which it is very valuable.

Most of the limitations discussed above are of greatly reduced significance when the assessment is formative rather than summative. In such circumstances the problem of negative marking for multiple choice questions is largely irrelevant as the point of the assessment is not to determine a mark but for the students to find out for themselves how well they know the material being tested. Likewise, the problem with students working back from the multiple choice options to the answer (as in the integration example in the section titled 'Multiple choice: Another problem') is less significant. If what is being assessed formatively is made clear then the main losers if the question is approached in the wrong way are the students themselves as they do not gain the information about their competence in integration that the assessment was designed to give.

Although the benefits of computer-aided assessment are most easily secured in formative assessment this does not mean that it cannot be used summatively. More care is needed to ensure that the assessment is robust and reliable, but this is possible. Summative use of computer-aided assessment for basic skills testing is well documented (see, for example, Beevers et al.[10] and Beevers et al.[11]).

One of the key issues now is whether to continue to seek to develop computer-aided assessment in mathematics to attempt to assess higher level skills or whether to accept that for the foreseeable future this can only be done through pen and paper assessments. Workers in this field must ask themselves some hard questions. It is undoubtedly intellectually challenging to seek creative and imaginative ways of using ICT to deliver assessment in mathematics. On the other hand, it is not very stimulating to mark a pile of student manuscripts. However, as time is a finite resource, we must ask if the returns from further advances in computer-aided assessment in mathematics are worth the resources that will have to be invested to achieve them. To put it bluntly, 'Will my students benefit more from me spending x hours marking their written work (in formative or summative assessment) than from me spending the same x hours experimenting (with uncertain outcome) in computer-aided assessment?'.

References

1. Quality Assurance Agency for Higher Education, (2000) *Handbook for academic review*, http://www.qaa.ac.uk/public/acrevhbook/intro.htm.
2. Bull, J. and McKenna, C. (2001) *Blueprint for Computer-assisted Assessment*, http://caacentre.lboro.ac.uk/bp/.
3. Ward, J. and Lawson, D. A. (2001) *Engineering Mathematics Testing and Assessment Pack*. Harlow, England: Pearson Education.
4. Hawkes, T. (1999) private communication.
5. Beevers, C. E. and Paterson, J. S. (2001) Automatic Assessment of Problem Solving Skills in Mathematics, http://ltsn.mathstore.ac.uk/articles/maths-caa-series/july2001/.

6. Strickland, P. (2001) How Should a Perfect Computer Aided Assessment Package in Mathematics Behave?, *MSOR Connections*, **1**(3), 49-51.
7. Assessing Mathematics in Higher Education, International Association for Educational Assessment (IAEA) Conference, September 1995.
8. McCabe, M., Heal, A. and White, A. (2001) Computer Assisted Assessment (CAA) of Proof = Proof of CAA: New approaches to computer assessment for higher level learning, http://ltsn.mathstore.ac.uk/articles/maths-caa-series/sep2001/index.shtml.
9. Lawson, D. A., Reed, J. and Tyrrell, S. E. (2002) A web-site for a mathematics support centre. In M. Borovcnik and H. Kautschitsch (Eds), *Technology in Mathematics Teaching*. Vienna, Austria: �bv&hpt, **2**, 149-152.
10. Beevers, C. E., McGuire, G. R., Stirling, G. and Wild, D. G. (1995) Assessing Mathematical Ability by Computer, *Computers in Education*, **25**, 123-132.
11. 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.

Acknowledgement

Duncan Lawson
School of Mathematical and Information Sciences
Coventry University
Coventry CV1 5FB
United Kingdom
d.lawson@coventry.ac.uk