Issues of partial credit in mathematical assessment by computer

C. E. Beevers, D. G. Wild, G. R. McGuire, D. J. Fiddes and M. A. Youngson

Department of Mathematics, Heriot-Watt University, email: c.e.beevers@hw.ac.uk

The CALM Project for Computer Aided Learning in Mathematics has operated at Heriot-Watt University since 1985. From the beginning CALM has featured assessment in its programs (Beevers, Cherry, Foster and McGuire, 1991), and enabled both students and teachers to view progress in formative assessment. The computer can play a role in at least four types of assessment: diagnostic, self-test, continuous and grading assessment. The TLTP project *Mathwise* employs the computer in three of these roles. In 1994 CALM reported on an educational experiment in which the computer was used for the first time to grade, in part, the learning of a large class of service mathematics students (Beevers, McGuire, Stirling and Wild ,1995), using the *Mathwise* assessment template. At that time the main issues identified were those of 'partial credit' and communication between the student and the computer. These educational points were addressed in the next phase of the CALM Project in which the commercial testing program *Interactive PastPapers* was developed. The main aim of this paper is to describe how *Interactive Past Papers* has been able to incorporate some approaches to partial credit which has helped to alleviate student worries on these issues. Background information on other features in *Interactive PastPapers* is also included to provide context for the discussion.

Introduction

The CALM Project for Computer Aided Learning in Mathematics started at Heriot-Watt University in 1985. In the first phase of the project CAL materials were created in calculus with each topic structured around Theory, Worked Examples, Motivating Examples and a Test. Students chose most readily to work through the Test section questions, welcoming the chance to assess their own progress. In addition, the teachers could view class progress. The weekly tests were designed from banks of questions with randomized parameters in each question. These questions prompted students for a mathematical answer and asked them to type in the response on one line using a style similar to computer languages such as Pascal. Students of engineering and science took only a short time to adjust to this approach. The routines developed in those early stages of CALM meant that testing could be more meaningful and did not rely on the more usual multiple-choice format favoured by so many computer projects. Over

the years 1989–1992 CALM developed techniques to trap predictable wrong answers and this form of self-testing proved to be a powerful learning aid for students. Nevertheless, some problems which are discussed in greater detail in later sections remained.

Despite these problems the assessment template which CALM created was used as a testing mechanism in collaborative projects such as the TLTP project, called *Mathwise*, and the SUMSMAN project. For further details see Harding and Quinney (1996), Beevers, Maciocia, Prince and Scott (1996), Beevers and Scott (1998) and Beevers, Bishop and Quinney (1998).

Meanwhile the CALM project focused on the production of a commercial project, *Interactive PastPapers* (IPP), aided by the award of a Higher Education Tercentenary prize from the Bank of Scotland. Based on the work of Wild (1995) the CALM Group were able to investigate the solutions to problems such as:

- providing partial credit where students receive recognition for partly-correct or idiosyncratically expressed answers;

- coping with difficulties students experience when they input answers.

See Wild, Beevers, Fiddes, McGuire, and Youngson (1997) and McGuire, Wild, Beevers, Fiddes, and Youngson (1998) for further details.

The features of IPP relating to these issues will be described later after a discussion of some of the educational issues raised by using computer assessment. The paper concludes with a summary of the way ahead in this area.

Educational discussion

Students of mathematics can be assessed in a number of ways. The computer can play a role in at least four types of assessment:

- diagnostic tests where the emphasis is on helping students discover their strengths and weaknesses;

- self-tests in which the emphasis is on rapid feedback and which can be used to pick out predictable wrong answers;

- continuous assessment in which students and teachers alike can see how mathematical topics are being absorbed;

- grading assessment in which the computer is used in the setting and marking of examinations.

Over the years there has been much work in the area of diagnostic testing with examples like *Diagnosis* (Appleby, Samuels and Treasure-Jones, 1997). Diagnostic testing using student profiling has been studied by Bridges and Hibberd (1994). This latter work is forming the basis of some current work on Web delivery of diagnostic tests.

Self-testing is a feature of the work of Harding, Lay, Moule, and Quinney (1995) in the Renaissance Project. CALM used this approach too in its pre-university mathematics courseware units, where predictable wrong answers were trapped and fast feedback helped students consolidate their learning. Self-tests are also a feature of the many *Mathwise* modules developed during the TLTP initiative.

Continuous assessment was the driving force in the first CALM Project with weekly tests helping the students assess their own progress through the material. Brunel, Portsmouth and Glasgow Caledonian Universities have used a similar approach with *QuestionMark* software and the CALMAT units (McCabe, 1995; Tabor, 1993).

Grading assessment by computer has been pioneered at Heriot-Watt University initially using the *Mathwise* assessment mechanism but latterly with the IPP assessment engine delivered over the Web. Napier University under the SUMSMAN Project are employing *Mathwise* to assess their students and this is starting in other Scottish universities (Ashworth, 1998).

The commercial implementation of IPP has a number of features that help with the different types of assessment described above. For example, diagnostic tests can be set up using the multiple-choice type of question. The marks recording service offered by IPP is then able to supply information on the strengths and weaknesses of individual students.

IPP has three modes of delivery:

1. Help mode where students can reveal answers if they are stuck on a particular part;

2. Practice mode in which the users are given visible feedback on the correctness of answers as they proceed through a question;

3. Examination mode in which the computer marks the question but no visible sign is shown on screen.

These three modes simulate self-test, continuous assessment and grading assessment respectively.

Questions can be chosen from banks of typical examples randomly, by topic or by particular question, thus providing different forms of testing at different times of the learning cycle. When a test has been chosen the student can browse the questions as in a conventionally written examination before moving on to answer a question. In the following section those aspects of IPP which deal with problems of partial credit and student input difficulties are considered.

Partial credit

When a student gets the same answer as the examiner to a mathematical question, assigning marks for the answer is simple both for a human and a computer. However, deciding how to assign marks to other answers for which there is not a one-to-one correspondence may be tackled differently by humans and computers.

Aspects of partial credit
A lack of confidence in the software's ability to assess answers has led to some student anxiety about whether their results in a computer examination accurately reflect their knowledge. Particular issues of greatest concern are:

*   the computer input of a mathematical answer being interpreted by the computer in a different manner from that which the student intended;

*   an answer being correct but in the wrong format;

*   recognition of a numerical approximation to the correct answer;

- provision for a student giving one of a pair of answers correctly;
- provision for a student answering a part of the question but not all of it.

The first of these has been tackled in IPP by the use of an Input Tool. This device shows the student how a one-line mathematical entry typed in by a student is being interpreted by the computer as the typing proceeds. For example, the student may wish the answer to be $\frac{1}{2x}$. The student types in 1/2x and the Input Tool window shows this as $\frac{x}{2}$.

This shows that the computer has interpreted this in a different way from that which the student intended, so it is possible to correct the input to 1/(2x) which appears in the Input Tool window as $\frac{1}{2x}$.

In addition, the Input Tool provides feedback on the syntax of expressions such as missing brackets and the misuse of operands.

Turning to the second of these items, answers which are close to the correct one but do not take the approved form can be awarded partial credit as in the following example. If the answer to a problem is 1/2 and the student types 0.5 then the message can be displayed:

Your answer is correct but in the wrong format, give your answer as a fraction – 50% partial credit.

IPP tries to capture good mathematical practice by looking for answers in a compact form. So, a maximum length can be specified and students awarded partial credit for their answer if it is close to the correct answer but not in the most compact form.

A similar approach works for the third of these topics in which partial credit is given for answers which contain a numerical approximation to the correct answer. If desired a warning message can be issued as in the following example. If the correct answer to a question is $\sqrt{2}$ the student who types in the numerical approximation 1.414 (which is correct to three decimal places) might receive the message:

Your answer should contain a square root – 75% partial credit.

From the last two examples it can be seen that the percentage of partial credit given can vary at the examiner's discretion.

Concerning the fourth of these topics, many questions in IPP require answers in the form of unordered and ordered linked pairs (or triples). Examples of such questions are:

What are the factors of $x^2 - 5x + 6$?

Find the co-ordinates of the y-intercept of the line $y = 3x + 7$.

The first question has answers $x - 2$ and $x - 3$ which can be given in any order, whereas the second question has the answer (0,7) in which the order is important. If one of these answers is correct then appropriate partial credit can be awarded.

A similar approach can deal with questions which require answers in scientific notation of the form $a*10^b$ with a and b as an ordered linked pair and algebraic fractions x/y with x and y again as an ordered pair.

The last issue is perhaps the most difficult to deal with. As an example of this type consider the question:

Find the derivative of $\cos(x^2)$

To do this question (whose correct answer is $-2x\sin(x^2)$), the student needs to know the correct rule of differentiation to apply and the derivative of two simpler functions. If a student were to answer $2x\sin(x^2)$ in which the correct rule has been identified and the derivative of one of the two functions has been correctly obtained, most human examiners would consider this answer worthy of some credit. CALM used a device in the self-assessment part of the pre-university CALM units in which predictable wrong answers were programmed into trapped errors so that such predictable answers could carry appropriate learning messages. However it could prove difficult to list all the possible wrong answers which the student came up with!

Key-steps and partial credit

IPP has questions which have been set using key-steps. These key-steps have been broken down into further sub-sets. It is up to the students to decide whether they wish to answer these sub-sets. A less confident student can tackle a question by asking for the sub-sets which would possibly enable some marks to be picked up.   However in such cases the student will take longer to answer the question but in Help or Practice mode this can aid the learning process. In Exam mode a student has to decide whether the extra time is worthwhile compared with the partial credit on offer in that question. This method of awarding partial credit has been steadily evolving over the last few years since the original experiment in 1994 was reported by Beevers, McGuire, Stirling and Wild (1995). Whatever the mode, this extra flexibility has proved popular with students and has been introduced after much student feedback on earlier versions of the CALM test.

IPP also provides an opportunity to deliver tests over the Web. So, in the winter term of 1997, two classes with in excess of 350 undergraduates were introduced to tests using IPP. Of the two classes using the IPP delivery over the Web the syllabus for the first class of over 200 students covered topics in differentiation including hyperbolic equations, parametric equations, the McLaurin and Taylor series and integration topics including standard integrals, application to differential equations, area under curves, methods of substitution and integration by parts. Two tests were planned with each one carrying a mark of 10 per cent. It was decided to give the students the use of the program in Practice mode so that they could see if they had errors as they went along.

The second class of about 150 science and engineering undergraduates took a course which covers topics in the theory and application of differentiation and aspects of complex numbers. In this second class only one computer test was set carrying a credit of 10 per cent of the module mark. Questions were set with one, two or three key steps. A student who could answer the question in the required number of steps could do so and move to the next question. However, each key step could be broken down into at most a further three sub-steps and the student who chose to tackle the question by answering every sub-step could score partial credit as the part answers are supplied. However, this had to be balanced in a timed summative test when the overall time to complete the examination is fixed. Clearly the design of the key-steps and sub-sets was critical in ensuring that students who could not tackle the entire question could get as much credit as they deserved for the parts they could do.

Towards an online examination for mathematics

It is now possible in a mathematical test to resolve the main difficulties of examining a student using the computer to set and mark the test. The use of random parameters in each question ensures that each student receives a test of similar standard, but cheating is impractical since neighbouring screens will carry different versions of the same question and hence will have different answers.

The design of the Input Tool has minimized the misinterpretation between student and computer. Moreover, the Input Tool carries an excellent syntax checker so that students are guided to form meaningful expressions. The Input Tool creates a dynamic display which provides immediate visual feedback on how the computer is understanding the student's input. As in many other testing situations if the student has practised with the use of the Input Tool then the questions asked on the day of the test hold little fear.

It is in the area of partial credit that most advances have been made. Through the design of questions and the introduction of key-steps, the less confident students can choose to take the question in smaller steps whereas the confident, careful student can move through a question more rapidly. Partial answers in the wrong format can carry some credit and some correct answers from a list can also be rewarded in part. Finally, the ability to deal with answers which are not correct can to some extent be done using key-steps and sub-steps.

So what remains to stop the creation of an on-line test remotely on demand? The assurance that the person sitting the test is the person they say they are! Such security issues remain the major stumbling-block to the setting and marking of a test remotely by computer. Some progress in dealing electronically with the issue of security is possible. Students can be screened by name, location and machine and, provided there is some validation of an individual's identity by a human checker, then an examination can be delivered. Answers can be recorded and updated at every input to prevent loss of results by electronic breakdown and results can be encrypted if necessary. It may be that the use of a video camera together with image-processing techniques can combine to remove even this restriction in the future (Daugman, 1997).

References

Appleby, J., Samuels P. and Treasure-Jones, T. (1997), 'Diagnosis: a knowledge based diagnostic test of basic mathematical skills', *Computers and Education*, 28, 113–33.

Ashworth, M. (1998), 'Maths resources on the Scottish MANs', *CTI Maths and Stats Newsletter*, 9, 32.

Beevers, C. E., and Scott, T. D. (1998), 'SUMSMAN-D collaboration between Scottish universities', *Proceedings of the International Conference on Technology and Collegiate Mathematics*, Chicago, 41–4.

Beevers, C. E., Bishop, P. and Quinney, D. (1998), 'Mathwise diagnostic testing and assessment', *CITech*, 191–205.

Beevers, C. E., Cherry, B. G., Foster, M. G. and McGuire, G. R. (1991), *Software Tools for Computer Aided Learning in Mathematics*, Ashgate Publishing Company.

Beevers, C. E., Maciocia, A., Prince, A. R. and Scott, T. D. (1996), 'Pooling mathematical resources', *Active Learning*, 5, 41–2.

Beevers, C. E., McGuire, G. R., Stirling, G. and Wild, D. G. (1995), 'Mathematical ability assessed by computer', *Computers and Education*, 25, 123–32.

Bridges, S. and Hibberd, S. (1994), 'Construction and implementation of a computer-based diagnostic test', *CTI Maths and Stats Newsletter*, 5.

Daugman, J. (1997), 'Face and gesture recognition: overview', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19 (7), 675–6.

Harding, R. D. and Quinney, D. (1996), 'Mathwise and the UK Mathematics Courseware Consortium', *Active Learning*, 4, 53–7.

Harding, R. D., Lay, S. W., Moule, H. and Quinney, D. (1995), 'A mathematical toolkit for interactive hypertext courseware: part of the mathematics experience with the Renaissance Project', *Computers and Education*, 24, 127.

McCabe, M. (1995), 'Designer software for mathematics assessment', *CTI Maths and Stats Newsletter*, 6, 11–16.

McGuire, G. R., Wild, D. G., Beevers, C. E., Fiddes, D. J. and Youngson, M. A. (1998), 'Computer aided learning in mathematics – Highers CALM Units and Interactive PastPapers', *Scottish Mathematical Council Journal*, 27, 53–7.

Tabor, J. (1993), 'Review of CALMAT', *CTI Maths and Stats Newsletter*, 4 (1), 14–18.

Wild, D. G. (1995), 'Computer assessment in mathematics', Ph.D. thesis, Heriot-Watt University.

Wild, D. G., Beevers, C. E., Fiddes, D. J., McGuire, G. R. and Youngson, M. A. (1997), *Interactive PastPapers for A-Level and Higher Mathematics*, Glasgow: Lander Educational Software.