The Maths, Stats
& OR Network

IT was twenty years ago today …

by

Cliff Beevers (c.e.beevers@hw.ac.uk)

Abstract

Since 1985 a team at Heriot-Watt University have delivered CAL/CAA software to undergraduates [1, 2]. Currently, the CAA package developed over two decades is offering formative testing to 25,000 school pupils on 60,000 courses through the SCHOLAR Programme [3] and summative assessment in an extension of PASS-IT [4,5] in two dozen Scottish secondary schools in a one-year trial. Formative online assessment within the university continues on a number of courses too. This article attempts to chronicle the successful ingredients of the Heriot-Watt CAA software, highlight the important milestones over the years and look forward to future work.

1. Introduction

Twenty years ago today, as Sergeant Pepper taught the band to play, the CALM Project for Computer Aided Learning in Mathematics began. Details of the work during the last twenty years can be found in [1, 2]. Over those intervening years, a number of lessons have been learnt which have enabled the successful delivery of computer based assessments to thousands of Service Mathematics students. Now, a new generation of school students are discovering for themselves the benefits of CAL and CAA through projects like SCHOLAR [3] and PASS-IT [4,5].
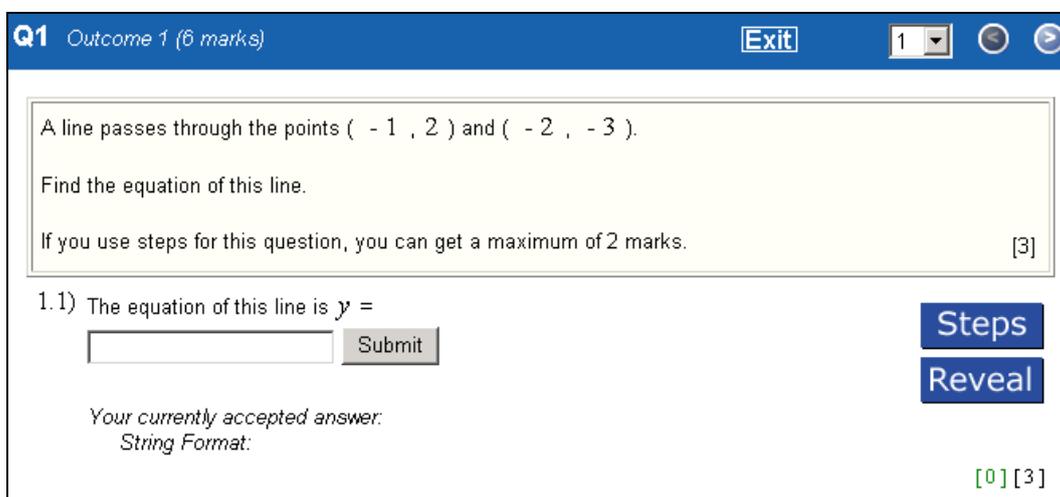
Essentially, CALM was a CAL package designed to engage students with elements of differential and integral calculus. CALM had a simple formula for success with each week's software package comprising core elements covering a review of theory, some worked exercises, motivating examples with a games flavour and a test section. The tests proved very popular with the students who saw them as a way of measuring their own progress, known as "assessment for learning" or "assessment as learning" or more formally formative assessment. The popularity of the test section by students together with the feature of reporting student work to the lecturer have remained central to the CAA delivery down the years [6,7].

One important ingredient of the very early CAA system was that students were asked to work out a mathematical expression as an answer to questions rather than select an answer from a set of options. Having the student enter their own constructed answer allowed questions of a more familiar form to be asked, rather than replacing them with a multiple choice format. It also enabled the student to submit their own answer, rather than using a process of elimination, or working backwards from the answer. In addition, the assessments usually contained randomised parameters inside the question text and answers. Randomised parameters enabled students to try the questions many times, receiving different versions of the question on which to practise their skills.
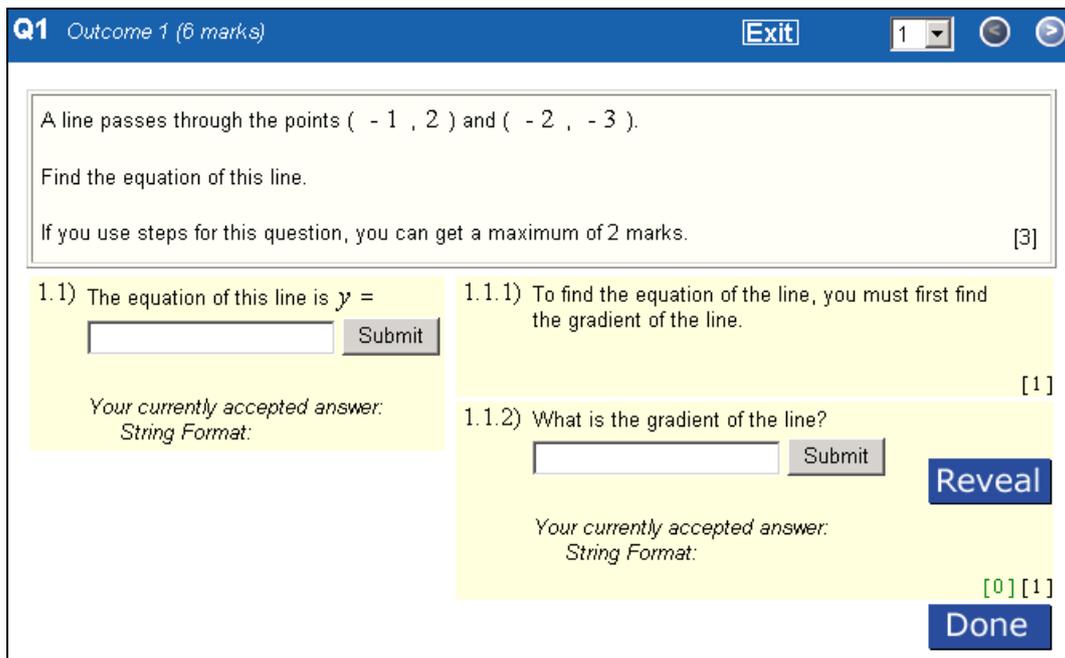
Implementing these features was not straight forward. Twenty years ago the CAS (Computer Algebra Systems) were in their infancy so a simple numerical approach was adopted by CALM to compare the correct and student answers at a number of points. This approach has not only stood the test of time but also has provided reliable marking throughout that time. This method

retains the feature that algebraically equivalent correct answers can be rewarded with a tick by the software. Its drawback is an inability to check for answers to questions like "Construct a quadratic with a turning point at x = 1." There are many such quadratics and an author in the old CALM system needed to provide a unique answer for the checking mechanism to handle the marking process. The new PASS-IT software relies on the same method of numerical comparison but through its integration with programs like JELSIM [8] more open-ended questions are now possible [9,10]. However, CAA remains a young science, more as an enhancement to the pen and paper approach, rather than a replacement measure. But, CAA can be very effective. It can ensure that the students are well grounded in the basics, sharpens their mental agility and breeds confidence as a course develops. CAA can also prepare students for tackling harder problems on paper at the end of the course.

Initially, CALM was designed as a formative CAA system, asking questions that required algebraic expressions as answers. As universities expanded and attracted students from more varied backgrounds the approach offered by CAL/CAA became even more attractive. Questions could be offered with different levels of help by, for example, using the approach of optional steps (Fig 1 and Fig 2).

Q1 *Outcome 1 (6 marks)* [Exit] 1

A line passes through the points (- 1 , 2) and (- 2 , - 3).

Find the equation of this line.

If you use steps for this question, you can get a maximum of 2 marks. [3]

1.1) The equation of this line is $y =$

Submit

Steps

Reveal

Your currently accepted answer:
String Format:

[0] [3]

Fig 1- An example of a question with optional steps where steps have not been revealed

Fig 2 - An example of a question with optional steps where steps have been revealed.

The good students can answer a question in one part whereas less gifted students may take several steps to reach the same conclusion. The idea of hiding these steps so as not to impede the more able student became a feature of the tests in the Mathwise modules [11]. Then, in the middle 1990s there was a drive to use some of the Mathwise tests summatively in some universities. Critiques of Mathwise assessment are provided by Beevers et al [12] and in this CAA series by Pitcher [13].

As these developments grew, it seemed sensible to consider some educational experiments to check that what was being proposed was both sensible and fair. This started a series of experiments which have continued as CAA has transferred from the tertiary to the secondary arena [14 - 18]. These research studies have investigated a number of issues including: what are the effects when a formative CAA system seeks to become summative? (see Beevers et al [14]), the effect of the delivery medium changing from pen and paper to computer and the effects of rewording and restructuring questions when converting paper questions to the computer (see Fiddes et al [15] and Ashton et al [17],).

Partial credit became an issue for students from these early experiments and gradually the role of steps evolved, not only within formative assessment, but also as a means of aiding the partial credit problem in summative testing (see McGuire et al [16] and Ashton et al [18]).

Alongside, and supporting the research, the development of the assessment system has progressed, always driven by the educational needs of the subjects and the people involved. Some development decisions are taken to deal with technical circumstances. For example, a real-time approach to the collection of assessment data was adopted in which students were asked to answer and submit answers as they moved through the test usually in multi-stage questions. Events such as answer submission, progression between questions and the use of steps are recorded on a remote server immediately as they occur, rather than collecting the data locally and bulk communicating this at the end of the assessment. This would have proved to be useful in the case of server failure, or local corruption of data, though this has rarely occurred. Other developments have been driven by the desire to ensure consistency, usability and accessibility. For instance, common standards and specifications, where they exist, have been employed to overcome issues such as the display of Mathematics on screen,

3

where MathML was introduced at an early stage to enable the creation of mathematical expressions on demand, including those involving random parameters. This has been invaluable in presenting the CAA accessibly. I am a registered blind teacher so this part of the work has been very satisfying. It was rewarded even more fully last year when the team took the mathematical tests to youngsters at the Royal Blind School in Edinburgh. There is much lip-service rhetoric around accessibility so it is an added pleasure to report true progress for such a disadvantaged group of pupils.

2. Some results

In the educational experiment described by Beevers et al [14], the three issues raised by students were how to avoid copying from adjacent computer users, the difficulty of entry of mathematical expressions into the computer and the problem of partial credit. Randomisation of questions, or parameters within a question template, can go some way to resolving the first difficulty though this in turn raises a fairness question: one student may receive all the tricky parameter values within a test, have more difficult arithmetic to handle and consequently fail the test. This criticism can usually be avoided by careful choice of the range of any random parameters within a question.

On the second issue CALM has held a consistent approach favouring the one-line input used by calculators with which many students are comfortable. One concession to convention was achieved early and retained throughout the years. This is implicit multiplication which is understood by the CAA software so users do not need to include the asterisk (*) (though students can choose to include it if they wish). The entry of mathematical expressions can still be problematic. For example, long expressions such as those which result from the differentiation of a quotient. Such inputs can cause a student to sigh audibly and at times require the endurance of a marathon runner. So, during the middle 1990s an Input Tool was constructed to help by displaying dynamically how the computer was interpreting the one-line entry of the mathematical expression. The Input Tool appeared in Mathwise [11] and in subsequent commercial offerings [19]. The move to Internet delivery in the late 1990s saw the demise of the dynamic input tool employed in Mathwise [11] and Interactive PastPapers [19] due to technical considerations. However, as PASS-IT progressed a static input tool was constructed which enabled students to view their submitted answers In rendered mathematical format. Judging from questionnaire returns it appears that this has gone a long way to resolving the mathematical expression entry problem. But, how to solve the partial credit issue?

Colleagues at Heriot-Watt turned their attention to the effect of the medium on the delivery of mathematical tests. Results in [15] indicated no statistically significant effect of the medium or re-wording of questions in Higher Mathematics (similar to AS level in England). It should be stressed that the conclusions drawn by Fiddes et al [15] were based on an assessment system that possesses the features which include algebraic expressions as answers and is able to provide optional steps. Further similar trials in other subjects at the upper end of secondary schools in the PASS-IT Project [17] confirm the earlier results on the effect of the medium and re-wording of questions. These results were investigated in a range of subjects and levels including Higher Chemistry, Higher French and Higher National Computing.

What gradually emerged for the role of steps in summative testing was the practice of awarding a penalty on those students who took them [16,18]. Such students would already be under some pressure in a timed test as taking steps creates more reading with more parts to answer. But, the research in the PASS-IT Project did indicate that such a penalty does not deter those who need to seek help while at the same time providing a fair test of mathematical ability, comparable with the conventional pen and paper tests. It should be stressed here that these trials [18] were aimed at assessments that measure minimum competencies only.

the Maths, Stats
& OR Network

As other workers in this field have found, some students need more help than questions that just ask "What's the answer". This is true for both formative and summative testing. Indeed, the role of steps in both formative and summative assessment is vital for some students. Formatively, it helps some students as they progress, and summatively, it can provide a way forward to the partial credit problem.

Another important lesson has been in the design of good CAA questions. Improvements in the types of questions and answering mechanisms have supported, and been driven by, the need to author appropriate and valid questions. However, the importance and strategies involved in good question authoring are not to be underestimated. This has been well-explained by Ashton and Youngson [20] in this CAA series. They describe, through a number of simple examples, the science of matching learning points to the steps or key parts of a question.

Throughout all this activity the information available to the teacher via the assessment system indicate that reports are essential. There is a distinct danger that without sensible feedback the teacher can be left out of the learning loop. Appropriate data collection, data reporting and choice of information display can enhance the teacher's role (the interested reader is directed to the articles [6,7] for more details).

3. Current work

As PASS-IT has come to a formal conclusion the work has progressed to the delivery of tests over the web to schools volunteering to take them in an online format. This work is at the level of Higher Mathematics at the moment in what are called the National Assessment Bank (NAB) tests. NAB tests are typically taken at three points within the course, are to measure minimum competencies and are pass/fail only. A student in Scotland must pass all three units in order to obtain the opportunity to sit the end of course assessment and hence gain a grade at Higher. The NAB tests provide solid building blocks on which to launch the conventional end of course assessment. In the traditional final assessment the topics of differentiation, integration, functions, trigonometry, vectors and logarithms are thoroughly tested.

As three typical examples from the NAB tests consider the questions below. Each is displayed to illustrate some of the issues raised above. The first instance of the question is its form on paper with question 1 taken from unit 1 on the equation of a straight line.

Question 1
A line passes through the points (-1,1) and (5,4). Find the equation of this line.

(3 marks)

The marks are awarded for knowing how to find the gradient, calculating it and then determining the equation of the line.

Here is the version as it is prepared for computer delivery with the random parameters a, b, c and d included.

Question 1
A line passes through the points (a, b) and (c, d). Find the equation of this line.

KP1: The equation of this line is y=?
Answer: (b-d)/(a-c)x + (ad-bc)/(a-c)

(1 mark)

> S1: To find the equation of a line, you must first find the gradient of the line.
>
> (Information only, 1 mark lost)
>
> S2: What is the gradient of this line?
> *Answer: (b-d)/(a-c)*
>
> (1 mark)

Note that KP1 denotes the prompt for key part 1 and S1, S2 are the optional steps as shown in Fig 1 and 2 above.

This question illustrates the role of optional steps in summative testing. When the question first appears on the screen this is the only prompt visible. A good student can answer the question at this point and score 3 marks. If a student cannot answer this question immediately then s/he can opt to press the "steps" button to reveal the two prompts denoted by S1 and S2. Notice that the first step provides the method but at the penalty of one lost mark. Thereafter, the student can score some marks by finding the gradient and then going on to complete the equation of the line in the key part thus gathering 2 out of the 3 marks and scoring some partial credit. In the recent PASS-IT extension trials in Scottish secondary schools the inability to find the gradient correctly was the most common fault in a sixteen question test which included elements of functions, the basics of differentiation and recurrence relations. The students had been able to practice with random values for the coordinates of the two points as set out above. Those that did practise consistently gained full marks when taking the NAB test for real. It should be stressed that the formative part of these summative trials is an invaluable part of the whole exercise.

Another example this time from Higher Mathematics unit 3 is as follows:

> **Question 2**
> A, B and C have coordinates (1, 4, 2), (2, 1, 6) and (4, -5, 14). Find the vectors **AC** and **BC** and hence show that the points A, B and C are collinear.
>
> (4 marks)

In this case the four marks are gained by finding the components of **AC** and **BC**, for interpreting the ratio between them and then for explaining why A, B and C are collinear. In all the five versions of this question in the paper NABs the three points are collinear.

With present technology it is not possible to ask a student to "Show …" that a particular result is true and mark it automatically. But, in an earlier trial another question which asks the student to "Show that the line L is tangent to a given circle C." had been problematic. The NAB tests on paper have five variants which the teacher can use but again in all five cases L is chosen to be tangent to the circle C. On translating this onto the computer one early idea was to ask the student to explain in words "How would you show that the line L is tangent to the circle C." This part of the question then had to be read and marked off-line by the teacher with the students answering this part of the question using free text. However, this type of explanation was well beyond many of the students. So, the question was changed into "Determine whether the line L is tangent to the circle C." as described by Ashton and Youngson [20]. Sometimes the random choice ensured that L was a tangent to C but other times that L cut the circle C at two distinct points. This makes the question a little more difficult than its paper counterpart but it does provide a good measure of the appropriate learning points to be tested.

Now, once again, in this case, all five examples of question 2 above in the paper NAB tests have 3 points which are collinear. So, this question does illustrate another important part of CAA in which a "Show that …" question can be neatly translated into a "Determine whether …" question. As the students practice the random parameters are chosen so that sometimes A, B and C are collinear but at other times they are not. Hence, KP3 and KP4 below have different answers depending on certain random parameters in this question.

The translation of this question into its electronic version for formative use is:

Question 2
A, B and C have coordinates (a, b, c), (d, e, f) and (p, q, r).

Find the components of the vectors **AC** and **BC**.
Hence determine whether the points A, B and C are collinear.

KP1: What is the vector **AC** in terms of i, j and k?
Answer: (p-a)i + (q-b)j + (r-c)k

(1 mark)

KP2: What is the vector **BC** in terms of i, j and k?
Answer: (p-d)i + q-e)j + (r-f)k

(1 mark)

KP3: If **AC** = s **BC**, what is the value of s, but if the equation is not true then enter s as zero?

(1.5 marks)

KP4: Are A, B, C collinear? Yes / No?

(0.5 marks)

Thirdly, a question on trigonometry from unit 3:

Question 3
Express 4 sin(x°) + cos(x°) in the form k sin (x + a)°where k >0
and 0° < a° <360°.

(5 marks)

The marks are awarded for knowing to expand cos(x-a), knowing its expansion, for comparing coefficients of cos(x) and sin(x), and for the interpretation of the comparison to find k and a.

The electronic version of the question has been translated as:

Question 3
Express bcos(x) + csin(x) in the form k * cos(x - a) where the arguments of the functions are measured in degrees, k is positive and 0 < a < 360.

(5 marks)

KP1: What is the exact value of k ?
Answer: sqrt(b^2+c^2)

(1 mark)

S1: You must first expand cos(x - a) .

(0.5 marks lost in this information step)

S2: What is the expansion of cos(x - a)?

○ cos(x)cos(a) + sin(x)sin(a)
○ cos(x)cos(a) - sin(x)sin(a)
○ cos(x)sin(a) + sin(x)cos(a)
○ cos(x)sin(a) - sin(x)cos(a)

(0.5 marks)

S3: What is the value of kcos(a) ?
Answer: b

(0.5 marks)

S4: What is the value of ksin(a)?
Answer: c

(0.5 marks)

KP2: What is the value of a in degrees (correct to 1 decimal place)?
*Answer: 180*arctan(c/b)/pi*

(1 mark)

S1: From the values of kcos(a) and ksin(a) it is possible
to calculate tan(a).

(0.5 marks lost in this information step)

S2: What is the value of tan(a)?
Answer: c/b

(0.5 marks)

This question illustrates the type of question students traditionally find difficult at this level. By the inclusion of 3 random parameters r, s and t which take values 0 or 1 it is possible to give the students a question in which the right hand side of the equation changes between cos(x+a), cos(x-a), sin(x+a) and sin(x-a) with the left hand side of the equation changing appropriately. This is achieved with an expression of the form r(scos(x-a)+(1-s)cos(x+a)) + (1-r)(tsin(x+a) + (1-t)sin(x-a)). With r = s = 1 and t = 0 or 1 the expansion is cos(x-a) is chosen. The breakdown of the question into steps like this provides some scaffolding for a student to work through a standard method providing valuable practice during formative assessment. During summative use of this question a good student can answer it in two key parts and score 3 marks and 2 marks respectively. If a student chooses to take steps at each key part then the maximum score would only be four marks out of the five on offer provided all the steps as well as both key parts are answered correctly.

Conclusions

More and more teachers are turning to CAA to help the consolidation of learning and in the measurement of that learning. It is important that further educational experiments are designed to ensure that this new form of testing is not a distortion. Specifications towards standards remain important but not crucial. It should be education driving technology and not vice versa. Good students will prosper in most situations. The main driver over the years for the Heriot-Watt team has been to widen participation, help the moderately able student succeed and aid the teacher deal with increasingly larger numbers of students at university.

So, what words of wisdom do I want to leave you with?

- Keep the questions fresh through randomisation;
- Seek answers using mathematical expressions wherever possible;
- Provide optional steps for those that need them;
- report all of this to both student and the teacher alike;
- Keep all of these new assessments accessible so that disabled students are not disadvantaged; and,
- Extend where possible to include animations, simulations and explorations so that higher order skills can also be tested electronically [9,10].

Finally, do all of this with a smile on your face so that the fun and satisfaction you derive from Mathematics is communicated to your students.

Acknowledgements

References

1. http://www.calm.hw.ac.uk/
2. H S Ashton, C E Beevers, C D Milligan, D K Schofield, R C Thomas and M A Youngson, *Moving Beyond Objective Testing in Online Assessment*, S L Howell and M Hricko (Editors), Online Assessment and Measurement: Case Studies from Higher Education, K-12 and Corporate, Information Science Publishing (ISBN: 1-59140-721-4
3. http://scholar.hw.ac.uk/
4. http://www.pass-it.org.uk/
5. H S Ashton, C E Beevers, A A Korabinski and M A Youngson, PASS-IT On, http://mathstore.ac.uk/articles/maths-caa-series/mar2005/
6. H S Ashton, C E Beevers, D K Schofield, M A Youngson, *Informative Reports - Experiences from the PASS-IT Project*, 8th International CAA Conference, July 2004
7. D K Schofield and H S Ashton, *Effective reporting for online assessment --- shedding light on student behaviour*, http://mathstore.ac.uk/articles/maths-caa-series/feb2005/
8. http://www.jelsim.org.uk
9. R Thomas, H S Ashton, B Austin, C E Beevers, D C Edwards, C Milligan, *Cost Effective use of Simulations in Online Assessment*, 9th International CAA Conference, Loughborough, July 2005
10. R Thomas, H S Ashton, *Use and Reuse of Simulations in Mathematical Online Assessment*, http://mathstore.ac.uk/articles/maths-caa-series/july2005/
11. http://www.bham.ac.uk/mathwise
12. C E Beevers, P Bishop and D A Quinney, Mathwise Testing and Assessment, Proc Eleventh International Conference on Technology and Collegiate Mathematics, New Orleans 35 - 39 (1999).
13. N Pitcher, Assessment with Mathwise --- Lessons, Limitations and Learning Curves, http://mathstore.ac.uk/articles/maths-caa-series/sep2002/
14. C E Beevers, G R McGuire, G Stirling & D G Wild, Mathematical ability assessed by computer, Computers and Education, 25, 123 - 132 (1995).
15. D J Fiddes, A A Korabinski, G R McGuire, M A Youngson and D McMillan, Does the mode of delivery affect Mathematics examination results? Alt-J 10, 60-69 (2002)

16. G R McGuire, M A Youngson, A A Korabinski and D McMillan, Partial credit in mathematics exams - a comparison of traditional and CAA exams, Proceedings 6th International CAA Conference, Loughborough University (2002)
17. H S Ashton, C E Beevers, A A Korabinski and M A Youngson, Investigating the Medium Effect in School Chemistry and College Computing National Examinations, British Journal of Educational Technology, 36(5), 771-787 (2005)
18. H S Ashton, C E Beevers, A A Korabinski and M A Youngson, Incorporating partial credit in computer-aided assessment of Mathematics in secondary education, British Journal of Educational Technology, To be published
19. C E Beevers, D J Fiddes, G R McGuire, D G Wild and M A Youngson, Interactive PastPapers, Lander Educational Software Glasgow (1997).
20. H S Ashton and M A Youngson, Creating questions for Automatic Assessment in Mathematics, http://mathstore.ac.uk/articles/maths-caa-series/feb2004/