Computer Assisted Assessment (CAA) of Proof = Proof of CAA: New Approaches to Computer Assessment for Higher Level Learning

by

Michael McCabe (email: michael.mccabe@port.ac.uk), Ann Heal (email: ann.heal@port.ac.uk), and Alison White (email: alison.white@port.ac.uk)

School of Computer Science and Mathematics, University of Portsmouth

Acknowledgement

This article was first presented at The Fifth International Conference on Technology in Mathematics Teaching, 6-9 August 2001 at the University of Klagenfurt in Austria, and will appear in the proceedings of the conference. For further relevant articles the interested reader is directed to http://www-sci.uni-klu.ac.at/ictmt5.

Introduction

We have used computer assisted assessment (CAA) successfully in support of maths teaching for 10 years, most commonly for first year university course units, where subject material is basic. This leads to the widespread belief that CAA is only appropriate for low-level learning. Nevertheless another reason why CAA is often used in first year course units is that student numbers tend to be greater, making the time spent on developing CAA more worthwhile. The assumption that CAA is inappropriate for all higher level learning is starting to be challenged (King, 2001) (Beevers, 2001).

Mathematical proof is a topic which students find difficult to grasp and involves a higher level of learning. We teach computer science undergraduates about the different methods of proof, as part of a second level maths for computing unit. The need for rigorous and precise arguments is important in, e.g. demonstrating the equivalence of two different computer codes, deriving results from graph theory applicable to computer networks, showing the equivalence of a recursive and non-recursive formula and proving the correctness of a recursive algorithm. The unit also lays the foundations for more advanced work on algorithms and formal methods.

We have used the Mathematics for Computing 2 unit as a test-bed for exploring the extent to which different forms of CAA can support the delivery of higher level learning. Initially our focus is on standard CAA, delivered interactively to each student by computer.

We follow this up by considering the contribution that group response systems, used in interactive classrooms, have had to play in promoting learning. Such group, or audience response systems have been described as 3rd generation CAA (McCabe, Heal and White, 2001) because of the extra dimension which they add to the student-teacher interaction.
Learning Outcomes

Higher level learning is commonly identified with the upper three levels of Bloom’s taxonomy of learning objectives. Bloom (1956) proposed 6 levels of cognitive learning, each associated with learning outcomes and a set of verbs.

Fig 1 - Bloom cognitive learning levels

The assessment of a learning outcome is achieved by a question or task based upon the corresponding verb. Each level depends upon achievement in the underlying levels.

A variety of alternatives, modifications and extensions to the Bloom classification scheme have been proposed. For example, Anderson and Krathwohl (2000) suggest independent classification of knowledge levels and a reversal of the upper two Bloom learning levels.

Fig 2 - Anderson and Krathwohl knowledge levels

This highlights the fact that the Bloom learning levels are neither sequential steps in problem solving nor measures of cognitive complexity. Furthermore their application in mathematics has often been reconsidered. For example, evaluation, the highest Bloom learning level, could be simply taken to mean checking that a numerical solution satisfies an equation or has a sensible value. Further lack of clarity arises from the mathematical uses of words such as evaluation (of a formula), analysis (of a function), solution (of an equation) and proof (of a theorem). Even the term applied mathematics is suggestive of “foundation thinking”.
Objective question design
Use of objective questions involving numeric, algebraic, text, multiple-choice and multiple response answers are commonly regarded as appropriate for foundation thinking, but incapable of assessing higher level learning. By adding advanced question types, from ordering and hotspot to drag and drop assembly and essay, it may be possible for higher levels of learning to be assessed. First of all, we consider how our questions have been designed from the commonly used techniques:

- conversion of traditional questions;
- use of templates; and
- use of verbs associated with learning outcomes.

One of us, Alison White, is a department-based educational technologist who helps teaching staff to deliver both formative and summative CAA in mathematics and computer science. Since teaching staff often have limited experience of objective testing, a common approach is for them to provide traditional questions for conversion. When the objective CAA questions are returned for approval, some are regarded as unsatisfactory, because they do not achieve what was intended. Another approach is to offer templates for generating similar objective questions, but this can often be viewed as limiting by academic staff. Nevertheless, we often use templates for designing summative test questions based upon formative test questions. This reduces authoring time and encourages students to practice on the formative tests. Learning outcomes are not usually specified explicitly.

Yet it is the failure to specify learning outcomes clearly in advance that can cause problems and lead to inappropriate questions. While learning outcomes are routinely specified for course units and for pieces of coursework, it is relatively unusual to see them on individual assessment questions. A critical test of CAA against equivalent means of assessment is whether it addresses the learning outcomes. If these are clearly stated for every question in advance, then it is much easier for staff to decide upon their suitability. If shown to students as well, they can identify more clearly what outcome is being assessed.

Objective questions for mathematics CAA
In mathematics there are other issues to consider. Learning levels could be based on how the subject is expressed in a hierarchy from words – numbers – tables – graphs – diagrams – symbols. The symbolic representation of algebra is then viewed as the highest level.

A common problem in identifying whether a learning outcome is higher level or not in Bloom-based taxonomies stems from the ambiguity of verbs. Apart from the obvious double-meaning of evaluation, mathematical tasks, such as calculate, solve and prove, cannot be pigeon-holed easily. Calculation and solution could be classified under application or analysis. Mathematical proof is even more tricky and is considered here.

Traditional exam questions on proof are time-consuming to mark, but CAA can provide an efficient alternative. The speed and accuracy of marking objective questions and the ability to give immediate feedback are among its obvious benefits.

First we looked at a set of 50 CAA questions designed to assess mathematical proof and delivered on-line using Question Mark Perception (http://meat.zen.port.ac.uk/mcom2/home.shtml - access restricted)
15 questions were used for formative assessment and 20 questions for summative assessment. A significant proportion of the summative questions used the formative questions as templates and every question had a clearly stated learning objective. We believe that

a) higher learning and knowledge levels were assessed. b) the questions were as effective as those traditionally delivered on paper. Yet even in the most simple cases the learning levels are not always apparent. For example, consider the basic question:

<table>
<thead>
<tr>
<th>Questions</th>
<th>Objective Types</th>
<th>Learning Level</th>
<th>Knowledge Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 basic concept of proof</td>
<td>multiple-choice, match</td>
<td>K,C</td>
<td>F</td>
</tr>
<tr>
<td>5 choice of proof method(s)</td>
<td>multiple-choice/response, match</td>
<td>K,C</td>
<td>F/P</td>
</tr>
<tr>
<td>11 write proof</td>
<td>order, fill-in-the-blanks, selection</td>
<td>App (A,S)</td>
<td>C</td>
</tr>
<tr>
<td>5 logic statements</td>
<td>multiple-choice/response</td>
<td>App</td>
<td>P</td>
</tr>
<tr>
<td>4 logical problems</td>
<td>multiple-choice/response</td>
<td>App</td>
<td>P</td>
</tr>
<tr>
<td>4 write algorithm</td>
<td>order, fill-in-the-blanks, selection</td>
<td>App (A,S)</td>
<td>C</td>
</tr>
<tr>
<td>4 apply algorithm</td>
<td>numeric, multiple-numeric, match</td>
<td>App</td>
<td>P</td>
</tr>
</tbody>
</table>

Fig 3 - Properties of questions for assessing mathematical proof

15 questions require students to ‘write’ a mathematical proof or a computer algorithm. They seem to be above the learning level of foundation thinking and are singled out for attention. A further issue arises because, although each question has a clearly stated learning objective, two or more questions taken together may assess a higher learning level. For example, one question can ask for the identification of appropriate proof methods for different mathematical statements (Fig 3), the next question can ask for a complete proof using an objective question (Figs 4 and 5).
Learning Objective - to be able to select an appropriate proof method

Choose the most appropriate proof method for each of the following statements:

2^n > n^2 for natural numbers 0 < n < 4
The square root of 3 is an irrational number
x^2 - 2x + 1 = 0 => x = 1
The sum of a rational number and an irrational number is an irrational number
For any connected graph, vertex connectivity ≤ edge connectivity
2^n > n^2 for all natural numbers n

---

Fig 4 - Selection Question Delivered On-Line using Question Mark Perception

---

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 multiplication 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.

---

Fig 5 - Order question delivered on-line using Question Mark Perception
Learning objective - be able to apply a proof by induction to a property of trees

Prove that any tree with n vertices has n-1 edges by placing the following steps in the correct numerical order:

Hence, the number of nodes for a tree with \(k+1\) vertices = \((k-1) + 1\)

Assume that a tree with \(k\) vertices has \(k-1\) edges

because if more than one edge was created there would be a cycle and if less than one edge was created there would be an unconnected node

= \((k+1) - 1\)

Addition of an extra node \(\Rightarrow\) exactly one extra edge is created

A tree with one node has no edges since there are no loops

By induction any tree with \(n\) vertices has \(n-1\) edges

= \(k\)

---

**Fig 6 - Fill-in-blanks (assembly) question delivered using Question Mark Designer**

The objective questions used for assessing proofs are of the following main types:

- order - place a set of statements or groups of statements in the correct order, where typically between 6 and 8 items are used, so that a longer proof would be grouped;
- fill-in-blanks(text) – complete several missing text or algebraic parts of a proof;
- fill-in-blanks (numeric) – complete several missing numerical parts of a proof;
- selection – choose from given items to complete several different parts of a proof; and,
- fill-in-blanks (assembly) – put together a proof from a set of fragments.

In the assembly question, each mathematical statement is broken into fragments, e.g. a letter for the first half and a number for the second. A proof is therefore answered as a fill-in-blanks text question in the form C4 A2 B3 D1. The example shown in Fig 6 has been investigated by analysing the responses of 150 students.

The student scores (Fig 7) are consistent with student abilities to generate a solution to the problem. Students who had no conceptual knowledge could not guess the proof and few students scored marks in the middle range. The spread of scores is exactly the outcome expected from a paper-based question requiring students to write out (create or synthesise) a complete proof.
Fig 7 - Student scores on the fill-in-blanks (assembly) proof question

When scores are considered in more detail (Fig 8) it can be seen how effective the question has been in awarding partial credit and hence in assessing lower levels of learning.

For example, some students could correctly place fragments or identify the first and second lines of the proof without being able to complete the proof. Similar results have been observed for other questions of this type.

A further point is that students often use rough paper to write out their answers, especially in ordering and assembly questions, before entering their answers on the computer. This may be seen as another indicator that the CAA is equivalent to paper-based assessment. Nevertheless there are other factors to consider. In particular, we recognise that question answering might be made more direct if graphical drag-and-drop components were used to implement ordering and assembly questions to avoid the indirect use of labels. The drawback is the extra authoring time required and the uncertain benefits to be gained. Indeed it may even be more beneficial to encourage students to write out their answers on paper.
Fig 8 - Details of scores on the fill-in-blanks (assembly) proof question

More advanced questions on proof include repeated statements. For example the question shown in Fig 9 states 15 propositions and 10 reasons which have to be assembled into a 16 line proof.
Other techniques include the identification of errors, although it is less straightforward to award partial credit and results are harder to analyse.

**Group response systems**

We have used a group (audience) response system, Varitronix PRS, to extend the use of CAA from computer laboratory to interactive classroom sessions (McCabe, Heal and White, 2001). Students enter their individual answers via a handset. Their responses are collected and a summary of results is displayed graphically for immediate discussion. The Varitronix system is limited to the entry of an integer (choice) between 0 and 9 together with an optional confidence level of high, medium or low. Other more sophisticated systems are available, but they generally lack the portability.

A group response system provides immediate feedback both to the lecturer and the class. This type of CAA delivery has been extremely valuable for revision classes where students can remain anonymous and enter their level of confidence in their answers. The lecturer gives public feedback to the whole class and the questions can later be attempted privately to obtain individual feedback. Figure 10 shows a question delivered as part of an on-line test being used in parallel with the group response system. In this example the limitation of PRS input to a single integer means that the question can only be asked in stages, e.g. which line corresponds to the first step in the proof. Multiple choice and digit numeric questions can be asked directly, but other questions need to be adapted.
Although this seems to be a severe restriction, a major benefit of using a group response system is that follow-on questions can adapted according to the responses given by the class. If students identify the first line of the proof successfully, then they can be asked for the second line. If they cannot, then they can be asked to identify the value of n given that \( nx+1 \) is an odd number and so on.

Group response questions are often different from standard objective test questions, in that they are specifically designed to encourage interaction. For this reason some question types, not normally used in traditional CAA, may become appropriate.

<table>
<thead>
<tr>
<th>Question type</th>
<th>Example</th>
<th>Uses</th>
</tr>
</thead>
<tbody>
<tr>
<td>tryout or trial</td>
<td>( 2 + 2 = ? )</td>
<td>check system, identify misuse</td>
</tr>
<tr>
<td>polling</td>
<td>Which proof is hardest?</td>
<td>identify student problems first</td>
</tr>
<tr>
<td>ambiguous</td>
<td>How far is London?</td>
<td>encourage student response</td>
</tr>
<tr>
<td>provocative</td>
<td>For what integer is ( n^2 &gt; 2n )?</td>
<td>encourage student response</td>
</tr>
<tr>
<td>indiscriminate</td>
<td>How many are correct?</td>
<td>seek preliminary feedback</td>
</tr>
<tr>
<td>ill-defined</td>
<td>Solve ( x + y = 2 )</td>
<td>encourage student response</td>
</tr>
<tr>
<td>step-wise</td>
<td>Stages of proof/algorithm</td>
<td>help answer longer questions</td>
</tr>
<tr>
<td>branching</td>
<td>Which part shall we do?</td>
<td>flexible delivery of material</td>
</tr>
<tr>
<td>evaluation</td>
<td>How interesting was …?</td>
<td>seek student feedback</td>
</tr>
</tbody>
</table>
One question could ask which proof method was felt to be most difficult, before tackling a specific example. Another question could ask how many statements from a given list could be proved by induction. That would be pointless to ask in a summative test, since students could either identify the wrong statements or simply guess. In a formative test it does not matter. The question could subsequently be made more precise or adapted according to the answers given. Questions can also mimic "asking the audience" as used in popular TV quiz shows, and add an element of fun. In short, questioning for use in routine teaching can be made more subjective, less precise and more adaptive. We believe that the use of a group response system can be used, not only to reinforce the use of standard CAA, but also as a valuable tool for motivating higher levels of student learning.

The final word can be given to students who have been asked a range of questions to evaluate our use of the group response system, by using the system itself. A typical response was in answer to the question: “Does the use of PRS allow more class participation in revision classes?” (1=strongly disagree 5=strongly agree). The responses shown in Fig 11 show the strong level of agreement.

![Fig 11 - Student feedback on group response system in revision classes](image)

Whether effective computer assisted assessment of mathematical proof provides proof of higher level learning may be open to debate. The effectiveness of group response systems in encouraging learning is less doubtful, but the extent to which it can be exploited is still being investigated.

**References**

http://www.caaconference.com/pastConferences/index.asp or direct at: