PROVIDING FEEDBACK TO STUDENTS' ASSIGNMENTS

C J Sangwin
LTSN Maths, Stats and OR Network, School of Mathematics and Statistics, University of Birmingham

Assessment drives learning, and one crucial part of the learning cycle is the feedback given by tutors to students' assignments. Accurately assessing students' work in a consistent and fair way is difficult. Furthermore, writing tailored feedback in formative assessment is very time consuming for staff, and hence expensive. This is a particular problem in higher education, where class sizes are measured in the hundreds. This paper discusses when and how computer algebra, within a computer aided assessment system, can take on this role.

INTRODUCTION

This paper is concerned with assessment in mathematics, and in particular with the use of computer aided assessment (CAA). There are at least four types of assessment: diagnostic, formative, summative and evaluative (the latter concerns institutions and curricula). Following Wiliam and Black (1996), these terms *"are not descriptions of kinds of assessment but rather of the* use *to which information arising from the assessments is put"*. Wiliam (1999) reports that in mathematics, *"formative assessment has the power to change the distribution of attainment. Good formative assessment appears to be disproportionately beneficial for lower attainers"*.

In this paper we consider only formative assessment, in which we provide feedback to learners designed to encourage improvement. The purpose of this research is to examine students' answers, which are entered into a CAA system as free text, to investigate when and how feedback may be constructed automatically. We emphasize that only free text, in contrast to multiple choice, responses are considered.

One acute problem with CAA is that one only has access to students' answers and there is usually little or no evidence of their *method*. Clearly using only students' answers to a question is problematic. We consider a framework comprising the four possibilities generated by appropriate or misconceived methods resulting in correct or incorrect answers. If a student uses an appropriate method which results in a correct answer one gives some positive feedback. If their appropriate method results in an incorrect answer, perhaps because of a technical slip, partial credit for correct properties of their answer, and feedback for incorrect properties, can be provided. They may also be given encouragement and a further opportunity to answer. The case of misconceived method is more difficult. An incorrect response, which most often arises from a misconceived method, can attract identical partial credit for correct properties of their answer, and feedback for incorrect properties. Without intervention and support it is doubtful if encouragement and a further opportunity to answer will significantly help the student, unless method hints are included in the feedback. Perhaps the most worrying case is when a misconceived method results in a *correct* answer. If we only base the response on this answer we must award full credit. The misconception is then left to fester dangerously, with the student having been rewarded.

Although this study takes place in higher education, the data presented below is taken from the beginning of a first year core calculus and algebra course. Hence, the material would not be out of place in A-level pure mathematics. This material was chosen so that the research might be applicable to the broadest range of mathematics educators.

METHOD

The CAA system used in this study is known as AIM, and is based on the computer algebra package Maple. Answers are entered as free text using a simple syntax, which are then tested objectively using computer algebra. The system tests the student's answer to check it has the required properties, rather than to see if it is the answer. In particular, AIM correctly assesses different algebraic forms of an answer. The system is also capable of giving partial credit and feedback. Another feature of this system is that in a multi-part question, should a student get the first part incorrect, the system can perform follow-on marking necessary to award credit for the subsequent parts. Designing such marking schemes is problematic, although this issue will not be address here. Students may also rectify their mistakes and repeatedly attempt any particular question, accruing a small penalty for each such incorrect response. Limited space precludes including screen shots of the working system. More details of using computer algebra based computer aided assessment can be found elsewhere, such as Strickland (2002); Sangwin (2003).

The study took place at the University of Birmingham over two years, where AIM is used routinely in the first year core calculus and algebra course, for diagnostic, formative and summative assessments. Formative assessments are set weekly to approximately 190 students, each comprising 10–20 questions. For each student, the system generates questions randomly, marks the work, and provides feedback to the student. Data is presented only for three questions, although we have similar data for each question we set using the CAA system described. To be concrete about methodology, let us consider our first question:

$$\text{Find} \quad \int x^{\frac{p}{q}}\,\mathrm{d}x. \qquad (1)$$

For each student two random integers were generated satisfying

$$p \in \{3, 5, 7, 11, 13\} \quad \text{and} \quad q \in \{2, \cdots, 12\}$$

with the extra condition that if $p = q$, then use $q + 1$ instead. The system tracks which random numbers have been assigned to each student and marks their responses accordingly. Data for this question will be presented in the results section below.

We believe that using such randomly generated questions reduces the problem of plagiarism, but more importantly for research purposes it allows one to check for any bias introduced to the data by particular choices of numerical values. In this case, the absence of a constant of integration was ignored, although being strict in requiring one is a possibility.

The system itself collates and presents the data, both aggregate data about individual questions, and the responses of individual students. The system groups data according to the values of p and q used. A typical response for one pair $(p, q) = (13, 10)$ being

```
QuestionNote = "Int(x^(13/10),x) = 10/23*x^(23/10)"
  Answer = "10/13*x^(13/10)" x 1
  Answer = "10/13*x^(23/10)" x 2
```

The `QuestionNote` gives information about which random numbers have been used in the form of the problem and the correct solution. The `Answer =` displays the answers given to this version of the question, together with the number of such responses, (ie `x 2`). In this case the system has been instructed to summarize only answers which did not get full marks.

Qualitative data pertaining to a student's method for answering a particular question is gathered using online surveys which take place immediately after a student has answered such a question. Experience of using interviews, paper based questionnaires and such online feedback suggests that the immediacy of the online format gives both reliable and rich data. Students also seem more willing to express themselves online than they do in writing. Interviews are time consuming, whereas the online form allows each student to be surveyed and automatically matches their responses with their question. Hence, the survey is not strictly anonymous since the researcher can match comments with particular answers – which is obviously advantageous. Once the test is complete the data is examined to identify patterns of incorrect responses, and methods. This is presented below.

RESULTS

Indefinite integration

The responses to this problem are taken from the 2003 cohort of 190 students. Each student was asked this question only once, and we consider only their first attempts, although repeat attempts are possible and encouraged. Some 34% of students made an error on their first attempt, and by examining these the mistakes shown below were identified.

$\frac{1}{n}x^{n-1}$	1	$\frac{1}{n}x^{n}$	1	$\frac{1}{n}x^{n+1}$	17
nx^{n-1}	0	nx^{n}	1	nx^{n+1}	2
$(n+1)x^{n-1}$	0	$(n+1)x^{n}$	2	$(n+1)x^{n+1}$	16
$\frac{1}{n-1}x^{n-1}$	1	$\frac{1}{n+1}x^{n}$	4	$\frac{1}{n+1}x^{n+1}$	na

Other	19
Syntax error	16

Forms and frequencies of students' incorrect answers.

By syntax error we mean results which are syntactically correct but which are not what the student *probably* meant. For example, a student's answer to $\int x^{\frac{5}{4}}\mathrm{d}x (= \frac{4}{9}x^{\frac{9}{4}})$ was `4/9*x^9/4` $= \frac{4}{9}x^9\frac{1}{4} = \frac{1}{9}x^9$. While this is incorrect, we know what the student meant. The other errors do not fit the above categories, although most of these could have been the answer to an integration problem, eg $\int x^{\frac{5}{6}}\mathrm{d}x = \frac{6}{7}x^{\frac{7}{6}}$.

It is straight forward to test for each of these errors, in an identical way to testing for the correct answer, and provide feedback to suggest how to correct a mistake. However, only three errors occur with a frequency which is significant. In addition to this, the CAA may operate on a particular answer such as $\int x^{\frac{5}{6}}\mathrm{d}x$ evaluated erroneously as $\frac{6}{7}x^{\frac{7}{6}}$ in order to automatically generate feedback such as the following.

> The derivative of your answer should be equal to the function that you were asked to integrate, which was: $x^{\frac{5}{6}}$. In fact, the derivative of your answer is $x^{\frac{7}{6}}$ so you must have done something wrong.

Below are error rates at a first attempt to solve (1) for different values of p and q.

p	1	3	5	7	11	13
%	28	30	38	38	26	39

q	2	3	4	5	6	7	8	8	10	11	12
%	53	33	28	40	46	60	38	22	44	18	6

Considering only p, we see no significant differences between values. However, different values of q have a marked effect on students error rates. Interestingly, there was no significant differences between $p < q$ (proper fraction) and $p > q$ (vulgar fraction), which might have been expected.

Quadratic equations

In this section we present results from the following two part question.

(i) Give an example of a quadratic, with roots at $x = p$ and $x = q$.
(ii) What is the gradient at $\frac{p+q}{2}$?

In this question the numbers p and q were randomly chosen for each student so that $p \in \{1, 2, 3\}$, $r \in \{1, 2, 3\}$ and $q := p + 2r$. Thus $p + q$ is even, and so the number given in part (ii) is an integer. Note, there is a family $\alpha(x-p)(x-q)$ of correct answers, for all $\alpha \neq 0$, and regardless of which α is chosen the answer to the second part is 0. This question was carefully designed so that it was possible to answer the second part correctly, without answering the first part. The students were asked in an attached questionnaire to explain briefly their method in answering these linked questions.

Of the 177 responses to this question, 80% were correct on the first try and a further 14% were correct on the second attempt. Only 2% of students did not correctly answer part (i) eventually. Similarly, of the 174 responses to part (ii), 71% were correct on the first attempt, with a further 13% correct on the second. However, 6% of students made 10 incorrect attempts or failed to correctly answer the question.

What is more interesting are the results of the feedback questionnaire. For the first part the vast majority of the students' concept of a quadratic is an expression in the unfactored form. The following is a typical student response.

> If the roots are a and b you multiply out $y = (x - a)(x - b)$ to get the quadratic y.

However none of the students demonstrated knowledge that the factored form was equally acceptable, required less calculation, and gave fewer opportunities for error.

The feedback for the second part was more revealing. Encouragingly, 86% of students *"differentiated and then subs in the given x value"*, which is is one very obvious strategy. Some 3% of the students misread the question and *"Used differentiation and put this equal to zero to find the turning points"*. Similarly, 3% used geometrical thinking and described their method as *"5 is the mid-point of the two roots, 2 and 8, therefore parabola is at minimum. Gradient must be 0"*. What the feedback questions also revealed, were a number of misconceptions. Some of these were only mild, and would be ideal starting points for formative feedback, such as

> roots at 2 an 4, its common sense that the curve would haveto make a veryu sharp turn between these two, an the obvious place for this turn would be exactly between the two points so 3. Now gradient at a turning point is always zero!

Other responses revealed much more serious misconceptions which nevertheless resulted in the correct answer: 6% of students responded that *"$x = 2$ is a vertical line so its gradient was zero"*. This is an example of a correct answer derived from a misconceived method which examining the answer alone will never reveal.

Odd and even functions

Students were asked to give three different examples of odd functions, which the computer algebra evaluated by comparing $f(x) + f(-x)$ with zero. The data, which is presented in Sangwin (2004), revealed that x^3 was significantly more frequent than x, and that $f(x) = 0$ was absent. This corroborates previous research such as Selden and Selden (1992) which demonstrated that such singular (sometimes referred to as trivial or degenerate) examples are often ignored by students. A more subtle feature of the data is that the majority of the coefficients which are not equal to 1 are odd. It appears that students are making everything odd, not simply the exponents, eg $3x^5$, $5x^7$ were typical responses. Students' concept image (see Tall and Vinner (1981)) appears to include only a subset of functions captured by the agreed concept definition. While not incorrect, they posses a deficient concept image.

Feedback to address this misconception, which modifies the student's answer and gives some positive and hopefully thought provoking feedback, has been implemented. Essentially this adds one to each coefficient in a student's answer, if a student provided an odd function with odd coefficients all greater than one. The staff member may use the system to automatically collate forms and frequencies of students' responses and present these together with additional interesting or important examples. This could be used for class discussion to help integrate the CAA into a coherent learning cycle.

CONCLUSION

Examining only students' answers reveals any bias introduced by particular choices of random parameters. While this information may not be relevant for design of formative assessment, demands for equity in summative assessment requires consideration of this.

In some cases, such as the question (1), it is possible to produce high quality tailored feedback automatically which is based only on properties of answers and common mistakes. In other cases, a correct answer may obscure a serious misconception. Therefore, mechanisms which attempt to distinguish between appropriate and misconceived methods need to be further developed, integrated into the CAA and deployed routinely. Some styles of questions reveal deficient concept images, and where these can be identified positive feedback may be used to encourage students to consider other approaches or aspects of the topic.

REFERENCES

Sangwin, C. J.: 2003, Assessing higher mathematical skills using computer algebra marking through AIM, *Proceedings of the Engineering Mathematics and Applications Conference (EMAC03, Sydney, Australia)*, pp. 229–234.

Sangwin, C. J.: 2004, Assessing mathematics automatically using computer algebra and the internet, *Teaching Mathematics and its Applications* **23**(1), 1–14.

Selden, A. and Selden, J.: 1992, Research perspectives on concepts of functions, *in* G. Harel and E. Dubinsky (eds), *The concept of function*, Vol. 25 of *Mathematical Association of America Notes*, Mathematical Association of America, pp. 1–16.

Strickland, N.: 2002, Alice interactive mathematics, *MSOR Connections* **2**(1), 27–30. http://ltsn.mathstore.ac.uk/newsletter/feb2002/pdf/aim.pdf (viewed December 2002).

Tall, D. O. and Vinner, S.: 1981, Concept image and concept definition in mathematics, with special reference to limits and continuity, *Educational Studies in Mathematics* **12**, 151–169.

Wiliam, D.: 1999, Formative assessment in mathematics: (1) rich questioning, *Equals: Mathematics and Special Educational Needs* **5**(2), 15–18.

Wiliam, D. and Black, P. J.: 1996, Meanings and consequences: a basis for distinguishing formative and summative functions of assessment?, *British Educational Research Journal* **22**(5), 537–548.