[O26] Computer based assessment with short free responses and tailored feedback

Sally Jordan
Centre for Open Learning of Mathematics, Science, Computing and Technology (COLMSCT), The Open University
s.e.jordan@open.ac.uk

Keywords: computer based assessment, free text responses, feedback

Abstract

This paper describes the development of interactive computer marked assignments, providing students with instantaneous, targeted and relatively detailed feedback. Most of the questions are not multiple choice, but wherever possible the feedback given to students is tailored to their response to the question. Students are allowed three attempts at each question, thus enabling them to learn from the feedback provided. Questions are currently being developed that require free-text answers of around a sentence in length. These questions are authored with the help of a sophisticated linguistically based authoring tool, and early student responses are used to refine the answer matching. Challenges of designing, authoring, refining and providing feedback on suitable questions are described, along with early student reaction and reflections on the opportunities provided by assessment of this type.

It is widely recognised that feedback on assessment tasks has an important part to play in underpinning student learning, encouraging engagement and promoting retention (see for example Yorke, 2001). Gibbs and Simpson (2004) articulated eleven conditions 'under which assessment supports students' learning', and seven of these conditions relate to the provision of feedback. Online assessment provides an opportunity to give virtually instantaneous feedback, thus 'feedback is provided quickly enough to be useful to students' (Gibbs and Simpson condition 6). However, providing feedback which is targeted to an individual student's specific misunderstandings is more of a challenge for computer based assessment. Multiple choice questions enable specific feedback to be provided in response to predefined responses, but these questions narrowly constrain how students may respond (Sukkarieh and Pulman, 2005) and concern has been expressed over their reliability, especially when used for summative purposes (Lawson, 2001).

Throughout the UK Open University (OU)'s thirty-five year history, assessment has been seen as an integral part of the learning process. Tutor marked assignments (TMAs), although usually graded and thus having a summative function, also provide an opportunity for a student's tutor to give detailed and personalised feedback and a summary of points to assist with subsequent assignments. Feedback and 'feed forward' at this level is considered particularly important for students who are studying at a distance and have limited opportunities to participate in face-to-face or electronic tutorials. The OU also has a longstanding commitment to the use of appropriate media, including computers, to support student learning. Computer based assessment, initially in the form of batch-marked multiple choice computer marked assignments (CMAs), has been used since the early days, and since the 1990s, interactive formative questions of several types
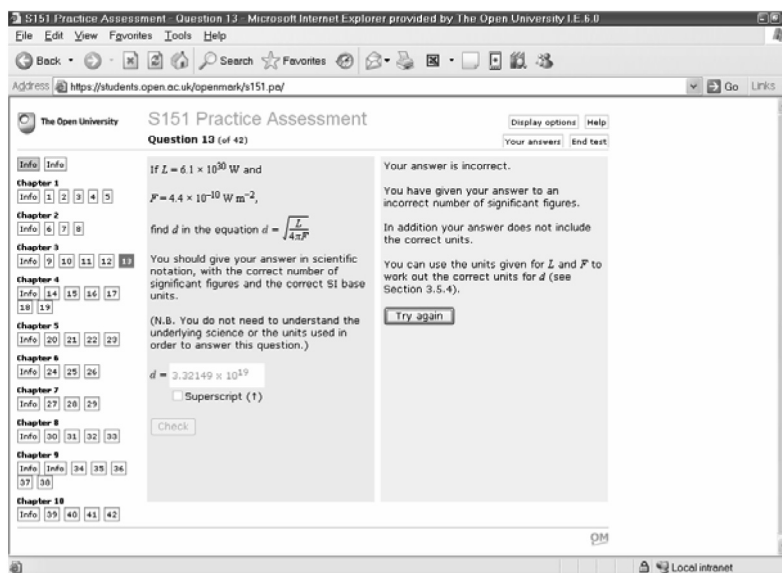
Figure 1: A question from the Maths for Science Practice Assessment, with targeted teaching feedback

(including drag and drop and numerical input) have been provided to Science Faculty students on CD-ROM and DVD-ROM.

In September 2002, a 10 CATS point course entitled 'Maths for Science' was presented to students for the first time. This course is studied over a relatively short time period (10 weeks – 3 months) and does not allow the opportunity for tutors to provide feedback to students via TMAs. However, the subject matter, and the fact that many of the students studying the course are lacking in mathematical confidence, meant that it was considered particularly important to provide rapid and meaningful feedback. It was therefore decided to use online interactive assessment for both formative and summative purposes. This enables students to be provided with instantaneous and targeted feedback on their responses to the questions, only about 25% of which are multiple choice (Jordan and Swithenby, 2005). **Figure 1** shows an example of the sort of teaching comment that may be provided, in this case in response to a student answer which is numerically correct but given to an unrealistic precision and without units. The feedback is targeted to the specific errors and includes a reference to the course material. The aim is to simulate 'a tutor at the student's elbow' (Ross *et al*, 2006), pointing out the student's error as specifically as possible, and providing a suggestion for how it might be corrected. The student is allowed three attempts at each question, with increasing amounts of teaching feedback provided after each attempt. The student is thus able to 'act upon the feedback to improve their work or their learning' (Gibbs and Simpson condition 6).

'Maths for Science' has now been studied by more than 7000 students and both the course and its assessment system have been well received. The technology underpinning the assessment system has been developed into the 'OpenMark' system, and OpenMark question types are being incorporated into the 'Quiz' function of the Moodle virtual learning environment. OpenMark is increasingly used for formative and diagnostic assessment by courses across the University.

Two new large population level 1 Science Faculty courses will be presented to students for the first time in 2007/2008, and both of these will use OpenMark questions in regular summative (but low stakes) iCMAs (interactive computer marked assignments). iCMAs will only form one part of the courses' assessment strategy (alongside conventional tutor marked assignments and, in one case, an end of course examination). The primary
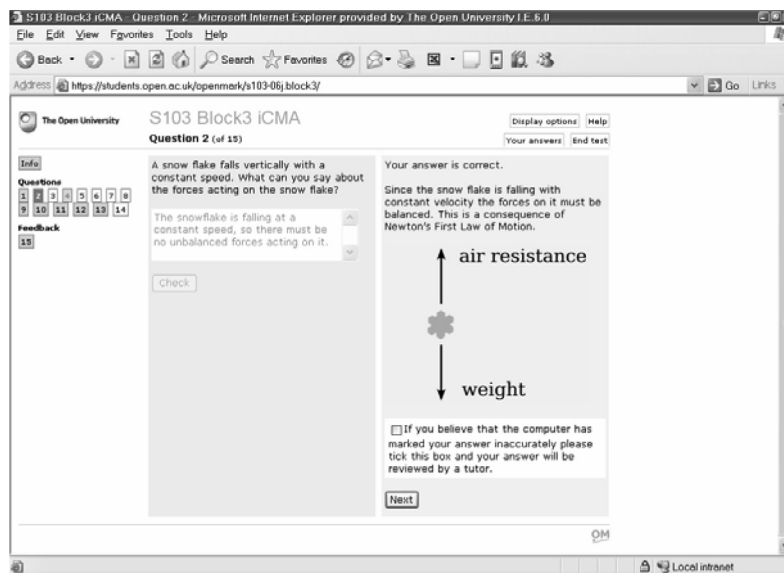
Figure 2: A typical (correct) student answer to a question requiring a short free-text answer

purpose of these assignments will be to encourage student engagement with the course material and to help students to pace their studies. However, the question types currently available mean that the tasks that can be assessed by OpenMark questions are limited, in particular questions requiring free-text answers are limited to those requiring numbers, symbols or answers to no more than a word or two.

In an attempt to explore other possible question types, a pilot study, jointly funded by the Centre for Open Learning of Mathematics, Science, Computing and Technology (COLMSCT) and the OU's VLE Project, is using an authoring tool supplied by Intelligent Assessment Technologies Ltd. (IAT) to write questions requiring free-text answers of up to around 20 words in length – typically a single sentence. The IAT authoring tool (Mitchell *et al*, 2003) is linguistically based, which means that an answer such as 'the dog bites the man' is recognised as being different to one such as 'the man bites the dog'. In the example shown in **Figure 2**, answers such as 'there are no unbalanced forces'. 'the resultant force is zero', 'the weight is equal and opposite to the air resistance', 'the forces are balanced' should all be marked as correct, but 'the forces are unbalanced', 'there are no forces' and the 'forces act in opposite directions' are incorrect or incomplete. The sophistication of the authoring tool means that these responses are indeed correctly matched, as are the vast majority of incorrectly spelt responses, responses in poorly written English, and many those which are expressed unexpected ways.

The Open University's pilot use of the IAT software has two major differences from that of most users of this software and other linguistically based short answer free-text systems (e.g. Sukkarieh and Pulman, 2005). Previous users have used student responses to paper-based questions in order to provide appropriate answer matching for the computer based version. The OU pilot recognises that 'real' student responses are essential as part of the developmental process, but we were reluctant to use responses gathered in one medium to inform the development of questions for a different medium – we felt that students might give different answers when working online and also that the optimal assessment tasks might be different. In order to obtain a batch of questions of sufficient reliability and robustness, questions are being developed with the assistance of students and staff on one presentation of the course 'Discovering Science' before being released to students on the next presentation of the same course (4 months later) for more extensive evaluation. Responses to the developmental versions of the questions are monitored regularly and the
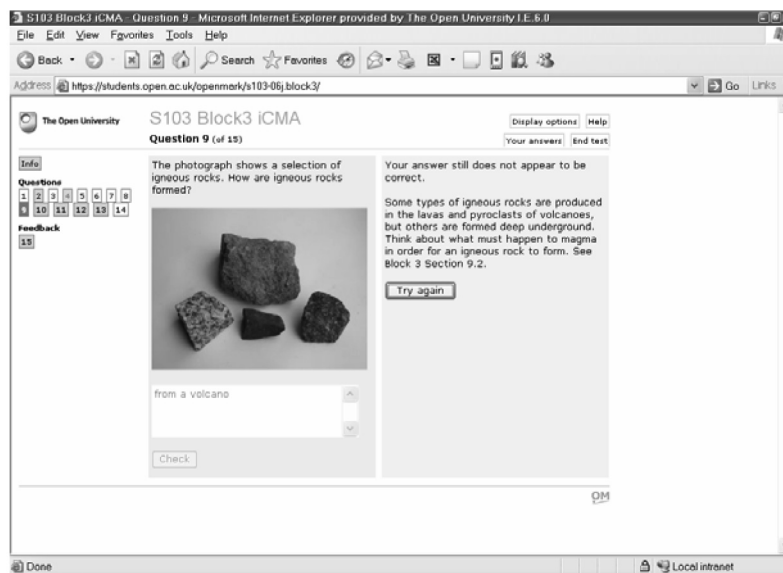
Figure 3: Targeted feedback on an incomplete student response to a question (Some igneous rocks are formed deep underground rather than being formed from lavas erupting from a volcano)

answer matching is amended whenever necessary (this can be done very quickly, whilst the system is live). The first batch of questions was released to students just before Christmas 2006 and by mid February 2007, the questions had been attempted by around 200 students, resulting in up to 277 responses per question for use in improving the answer matching. It is worth noting that, even in development, the questions have been very well received by students, with words such as 'fun' and 'addictive' being used quite commonly, and more than 75% of students reporting that they enjoyed answering the questions and found the feedback useful.

The second novel focus of the Open University's use of short answer free-text questions is the emphasis being placed on the provision of instantaneous teaching feedback. The IAT questions are currently being presented online to students via the OpenMark system, so a decision was made to replicate OpenMark's three stage feedback. At present the system attempts to recognise answers as correct or specifically or generally incorrect, and then to provide pre-written feedback including (as for conventional OpenMark questions) a reference to the relevant course material. Generic feedback is provided from OpenMark, but feedback targeted to specific student misunderstandings is generated within the IAT system. So, in **Figure 3**, the feedback in response to a student answer of 'from a volcano' (which is not considered a sufficiently complete response) has been generated from within the IAT system. Unfortunately the software used until February 2007 did not allow targeted feedback on many specifically incorrect student responses, and incomplete answers presented a particular difficulty. In the case of the question shown in **Figure 3**, it was not possible to provide targeted feedback on an incomplete answer such as 'they are formed from magma' because of interference with the correct answer 'they are formed from magma which has cooled and solidified'. However a new release of the software was supplied in February 2007 to enable the provision of targeted feedback whenever appropriate. For the present this feedback will continue to be written at the authoring stage, though it is recognised that there are other options for effective feedback on answers of this sort, for example letting students see which part of their answer 'matched'. A later stage of the project will compare the current system, based on computational linguistics, with a simple 'bag of words' system. This is likely to be less accurate in its marking, but may lend itself to different question types (including, paradoxically, those requiring longer answers) and the provision of different types of feedback.

The Open University's trial of online assessment questions requiring short free-text answers is in its very early stages, with four batches of questions released to students in their developmental phase, but only two of these batches released to students on the subsequent presentation of the course. An investigation has just begun into student perceptions of the questions and their use of the feedback provided. In addition, the effectiveness of the computer marking system will be compared against human markers.

Each time a student completes a question, they are provided with an opportunity to report if they think that the computer has marked their answer inaccurately. Most of the times this box is checked the student response has actually been marked correctly (and the student answer is usually wrong!). Questions which are relatively 'open' tend to generate the most uses of the 'inaccurately marked' box, perhaps because students fail to appreciate that they need to provide a precise answer (for example, if they are asked to compare intrusive and extrusive igneous rocks, they need to say *which* has bigger crystals). Framing appropriate questions, especially questions which assess understanding rather than simply recall, is, perhaps, a greater challenge for the question authors than using the authoring tool to generate appropriate answer matching. The time taken to write suitable questions and answer matching, and to modify this in the light of student responses, should not be underestimated. Using the IAT authoring tool to write questions of this type does not require any special linguistic abilities but it does require a logical approach to the authoring process.

The Open University's use of assessment of this sort is currently entirely formative; an 'add on' to a well established course. However, provided the questions prove sufficiently robust and student reactions are favourable, it is hoped to integrate a number of questions requiring short free-text responses into the regular summative assignments on the new level 1 course 'Exploring Science' from February 2008.

Monitoring student responses to interactive online assessment questions, essential in the development of sophisticated questions of the type described in this paper, can also be used to provide valuable feedback to academics about student misunderstandings. Analysis of student responses to the conventional OpenMark questions in summative 'Maths for Science' assignments has provided valuable insight into the mathematical misunderstandings of adult distance-learning science students (Jordan, in press). Even in the developmental phase, student responses to the short answer free-text questions has provided similar insight. For example, in answer to the question 'What does an object's velocity tell you that its speed does not?', a small but significant number of students gave the answer 'It tells you about a change in direction'. This answer was initially marked as correct (it is similar to the correct answer 'It tells you the direction in which the object is moving') but it is *not* a correct answer, and the source of the misunderstanding was tracked to an unfortunate section heading in the printed text. In both this case and the more systematic analysis of 'Maths for Science' questions, increasing understanding of student misconceptions has enabled changes to be made both to the assessment questions and to subsequent Open University teaching.

Acknowledgements

References

Gibbs, G. and Simpson, C. (2004) Conditions under which assessment supports students' learning. *Learning and Teaching in Higher Education*, **1**, 3-31.

Jordan, S. E. and Swithenby, S. J. (2005) Online summative assessment with feedback as an aid to effective learning at a distance. In *Proceedings of the 2004 12th International Improving Student learning Symposium, Birmingham*, ed Rust, C., pp 480-485. Oxford, UK: Oxford Centre for Staff and Learning Development.

Jordan, S. E. (in press) The mathematical misconceptions of adult distance-learning science students. *CETL-MSOR Conference, Loughborough, 11-12 September 2006.*

Lawson, D. (2001) Computer assisted assessment (CAA) in relation to learning outcomes. **http://ltsn.mathstore.ac.uk/articles/maths-caa-series/oct2001/index.shtml** (accessed 26th February 2007)

Mitchell, T, Aldridge, N, Williamson, W. and Broomhead, P. (2003) Computer based testing of medical knowledge. *10th Computer Assisted Assessment Conference, Loughborough, July 2003*.

Ross, S. M, Jordan, S. E and Butcher, P. G. (2006) Online instantaneous and targeted feedback for remote learners. In *Innovative assessment in Higher Education*, ed Bryan, C. and Clegg, K.V., pp 123-131. London, UK: Routledge.

Sukkarieh, J. Z. and Pulman, S. G. (2005) Information extraction and machine learning: auto-marking short free text responses to science questions. *Proceedings of the 12th International Conference on Artifical Intelligence in Education.*

Yorke, M. (2001) Formative assessment and its relevance to retention. *Higher Education Research and Development*, **20**(2), 115-126.