The definitive version was published in Open Learning: The Journal of Open, Distance and e-Learning, Volume 26 Issue 2, June 2011.
doi:10.1080/02680513.2011.567754
(http://dx.doi.org/10.1080/02680513.2011.567754)

Using interactive computer-based assessment to support beginning distance learners of science

Sally Jordan
The Open University in the East of England, Cambridge, UK
Email: s.e.jordan@open.ac.uk

Abstract
Interactive computer-marked assignments (iCMAs) are in use on a range of Science Faculty level 1 modules at the UK Open University, alongside tutor-marked assignments, for summative and formative purposes. They are also used diagnostically to help prospective students to make an appropriate choice of starting point. iCMAs have been very well received by students, but minor problems can easily lead to a loss of confidence. Students engage more with the questions when they carry some weighting and most students feel that their mark for iCMA questions should count towards their overall course score. Two case studies are presented, showing how evaluation has led to further improvements at both the question and whole assignment level.

Keywords
computer-aided assessment; science; diagnostic quizzes

Introduction
Whether we like it or not, assessment has a profound impact on learning. Assessment can define a 'hidden curriculum' (Snyder, 1971) or become 'the tail that wags the dog' (Dysthe, 2008, p.17). Boud (1995) has pointed out that whilst students may be able to escape the effects of poor teaching, they cannot escape the effects of poor assessment.

Reviews of the literature (e.g. Black & Wiliam, 1998; Gibbs & Simpson, 2004) have identified conditions under which assessment appears to support learning, and a number of frameworks have been devised for use by practitioners in developing and auditing their assessment practice (e.g. Gibbs & Simpson, 2004; Nicol & McFarlane-Dick, 2006). These frameworks share common themes centred on the importance of student engagement and the influence of feedback on learning. Gibbs (2010, p.163) recognises feedback as having 'more leverage to improve learning than any other variable', but in order to be effective, feedback needs to be more than a one-way transmission of information from teacher to learner. The effectiveness of feedback is enhanced by the prior engagement of the student with the task

(Hattie & Timperley, 2007), even if the student makes errors (Kornell et al, 2009). The effectiveness is also enhanced when the student acts on the feedback received, enabling them to reduce the gap between their previous understanding and that of the teacher (Sadler, 1989). In these ways feedback becomes a process not a transaction (Havnes & McDowell, 2008, p.118).

The broadest definition of e-assessment (adapted from Sim et al, 2004, and Whitelock & Brasher, 2006) encompasses the use of computers for any assessment-related activity, thus it might include the electronic submission of tutor-marked assignments, the marking of student engagement with a tutor group forum or the compilation and grading of an e-portfolio. However the work described in more detail below relates to the online delivery and automatic marking of questions, with the provision of immediate feedback, enabling Gibbs & Simpson's Condition 6 ('Feedback is timely in that it is received by students while it still matters to them and in time for them to pay attention to further learning or receive further assistance') to be met very easily. Students can be offered repeated attempts to practise (Bull & Danson, 2004) and the regular use of e-assessment can help them to monitor their progress and pace their studies. Thus in a variety of ways, e-assessment is acting as a virtual 'tutor at the student's elbow' (Ross et al, 2006, p.125), which is particularly important in a distance-learning environment.

Not only can feedback be provided instantaneously, it can also be provided without requiring tutors to mark and comment on individual scripts. Thus there are potential savings of effort and resource. E-assessment is available to students wherever they are in the world and is the natural partner to e-learning (Mackenzie, 2003), providing an alignment of teaching and assessment modes (Gipps, 2005). There is much optimism about the possibilities offered (e.g. Whitelock & Brasher, 2006) but also some anxiety that inappropriate use of e-assessment will have a negative effect on learning, in particular encouraging a surface approach (Scouller & Prosser, 1994; Gibbs, 2010). Some of these concerns appear to be founded on the belief that e-assessment is limited to the indiscriminate use of multiple-choice questions. However, even with the increased availability of more sophisticated e-assessment items, it remains important that online assessment is not seen as a panacea, but rather used only when appropriate and in balance with other types of assessment (Mackenzie, 2004). Most importantly, the primary driver should be not financial gain nor the use of new technologies for the sake of doing so, but rather a desire to improve student learning.

Open University level 1 science modules and their assessment strategies

For much of its 40-year history, students at the UK Open University (OU) who wanted to study for a science degree were encouraged to commence their studies with successive generations of the *Science foundation course*. The current version of this is *S104 Exploring science*, a 60 CATS point module at OU level 1 (Framework for Higher Education Qualifications level 4), which introduces students to biology, chemistry, physics, astronomy, Earth and environmental science, and to appropriate mathematical and study skills that they will need for subsequent study. S104 is offered in two nine-month presentations each year and is studied by around 4000 students per annum. S104 is complemented by the 10-point residential school module *SXR103 Practising science*, which includes a week at residential school in which students gain hands-on experience in the laboratory and in the field. However S104 is no longer the appropriate or favoured starting point for all intending science students. An increasing number of students do not have the basic scientific and mathematical skills that are assumed by S104, and these people are encouraged to start their studies with *S154 Science Starts Here*, a 10-point ‘gateway’ module which runs for 10 weeks. Those who want to concentrate on health sciences, are advised to start with *SDK125 Introducing health sciences*, a 30-point module. In addition, a range of short 10-point and 15-point modules are offered, all at level 1, with subjects as diverse as *Diabetes care*, *Elements of forensic science* and *Understanding the weather*. Many of these are chosen by new students, either because the subject matter is of particular interest to them or because they want to get a taste of Open University study before committing themselves to a longer module.

On the longer modules, students are supported by tutors who mark and comment on tutor-marked assignments (TMAs), give tutorials (increasingly using a range of synchronous and asynchronous communication technologies) and offer one-to-one support when necessary. TMAs may assess course components that are delivered electronically (e.g. a virtual fieldtrip on a DVD) as well as online activities involving information literacy and group working skills (Kirkwood & Price, 2008). TMAs were initially submitted on paper through the postal system, and returned from tutor to student via the OU headquarters, again by post. TMAs are now more likely to be submitted electronically via the eTMA file-handling system. This enables TMAs to be submitted and returned with a minimum of delay, saves printing and postage costs and facilitates the use of the plagiarism detection software CopyCatch and Turnitin®. Despite problems with inputting and commenting on diagrams, graphs and symbolic notation (Freake, 2008), eTMAs have been very well received by students. S104 offers students the choice of electronic or paper submission; in the first presentation of the module in 2008, 80-85% of students submitted their TMAs electronically but by 2010 this figure had risen to approximately 95%.

Most of the modules described above also use interactive computer-marked assignments (iCMAs). For example, in S104, seven TMAs are complemented by eight 10-question iCMAs and the synoptic script-marked end-of-course assessment is complemented by a longer 25-question iCMA. Overall, iCMAs account for 25% of the course score so each of the 10-question iCMAs contributes around 2% to the final score. The primary aim of the iCMAs is to provide students with instantaneous feedback; they carry a small summative weighting simply to encourage students to engage with them. For other modules, similar summative iCMAs and TMAs are accompanied by purely formative ‘practice’ iCMAs, which students can attempt as many times as they would like to. The SXR103 iCMA is purely formative, designed to help students to check their understanding of the pre-residential school study material before they attend the school.

Students use their own computer (or an internet cafe or library) to complete iCMAs. In summative use, iCMAs are typically made available for a period of a few weeks while students are studying the relevant material, with a hard cut-off date to help students keep up to date, but no limit to the amount of time spent actually working on the iCMA within that period. Formative-only iCMAs are usually available for the duration of the module.

Beyond multiple choice

The Open University uses the Moodle virtual learning environment and is global maintainer of the Moodle quiz engine. Added functionality, both in terms of the question types that can be used and the way in which iCMAs can be put together, is provided by the use of the 'OpenMark' system, developed at the Open University but now open source (Butcher, 2008). Whether the questions are written in OpenMark or Moodle and whether they are summative or purely formative, the Moodle Gradebook enables students to monitor their own progress, encouraging sustainable self-assessment practices (Boud, 2000). Tutors can monitor their students' progress and offer extra support if appropriate.

The simple question shown in Figure 1 illustrates some of the features of OpenMark. Even in summative use, students are given several (usually three) attempts at each question, with increasing feedback provided in response to incorrect responses and a correct answer provided when the student has either used all their attempts or got the question right. An incorrect response at the first attempt typically results in the simple feedback 'Your answer is incorrect', to give students the opportunity to work out their error for themselves. After a second incorrect attempt, there is more detailed feedback, usually with a reference to the relevant course material. Wherever possible the feedback is tailored to the error that the student has made.

Most OpenMark questions exist in several variants, for example there are five very similar variants of the question shown in Figure 1; another variant is 'Work out $\frac{1}{5}+\frac{1}{7}$, entering your answer as a fraction using the boxes provided.' In summative use, multiple variants are used as an anti-plagiarism device, with different students receiving different versions of each question. In formative-only use, the different variants provide extra opportunities for practice. Whereas in summative use it is important that the different variants are as similar as possible, in formative-only use some variability (e.g. asking students to subtract one fraction from another rather than to add two fractions) can be considered to be advantageous.

OpenMark supports a range of question types including multiple-choice (the selection of one option from several), multiple-response (the selection of multiple options), drag-and-drop, and hotspot (where a location on the screen has to be specified, useful for example in practising graph-plotting skills). In addition to the free-text entry of simple numbers, as shown in Figure 1, OpenMark also supports the entry of letters and words, whilst a superscript/subscript function enables the entry of scientific notation and chemical formulae. Different components of an answer can be checked for marking and feedback e.g. is the numerical answer correct?, is the stated precision of the answer reasonable?, are the units correct?

It is possible to use simple multiple-choice questions creatively in order to create e-assignments which assess higher order skills, motivate students to engage, and encourage students to self-reflect (for example, Nix & Wyllie, in press, used a confidence indicator tool and a learning log). However questions in which students construct their own answer, rather than choosing from pre-defined options, are considered to require deeper cognitive processing (Nicol, 2007). A project funded by the Centre for the Open Learning of Mathematics, Computing, Science and Technology (Jordan & Mitchell, 2009) has investigated the use of questions in which students are required to give their answer as a free-text sentence of up to 20 words, and twenty-five questions of this type are now in regular use. The answer matching for these questions was initially written using commercial linguistically-based software, but following the surprising finding that equally accurate answer matching could be obtained using much simpler pattern matching software (Butcher & Jordan, 2010), these questions are now offered to students from within OpenMark.

Are you ready?

The Open University's open entry policy means that undergraduate modules have no mandatory entry requirements and students commencing their study of science come with a wide range of previous qualifications, and none. The plethora of level 1 science modules, whilst facilitating student choice, could also easily lead to confusion. Diagnostic quizzes have an important role, alongside educational guidance, in helping prospective students to make appropriate choices and to only study modules for which they are adequately prepared. Since many OU students study alongside employment and/or caring responsibilities, it is also important that enquirers are discouraged from registering for modules which they have insufficient time to study.

For many years, the Science Faculty has produced a range of self-diagnostic 'Are you ready for?' materials. These were initially printed and more recently have been made available as a series of pdf files on the faculty website. It is difficult to monitor the use of the printed and pdf materials, but it is known that prospective students sometimes look at the worked solutions to the diagnostic questions before attempting the questions for themselves, and then assume that they would have been able to answer the questions just because the solutions make some sort of sense to them. It can also be tempting for prospective students to over-estimate the amount of time that they will be able to make available for their studies.

An integrated online diagnostic quiz, 'Are you ready for level 1 science?', was produced in 2007, to direct students towards the most appropriate starting point on the basis of their ability to answer basic mathematics and English questions and on the amount of time that they identified as having available for study. 'Are you ready for level 1 science?' was produced as three interlinked OpenMark assignments, though an attempt was made to hide the complexity from users. Students started with an introductory quiz which assessed the amount of time that they had available, their preference for studying over nine-months or a shorter period, and their ease of access to a computer and the internet. They were then directed as appropriate to one of two quizzes to assess their preparedness for (i) S104 or (ii) the other level 1 modules.

Every effort was made to make the questions as interactive as possible. For example, rather than investigating the amount of time that students had available for study by asking whether they had, say, 5-10 hours, 11-15 hours or 16-20 hours available each week for studying, students were asked to enter the actual time they could find in the morning, afternoon and evening of a typical Monday, and so on. Their total time per week was then computed and used to give appropriate feedback. Many of the questions for the diagnostic quiz were variants of the questions in formative and summative iCMAs. This re-use of question templates enabled the diagnostic quiz to be produced at much reduced cost. Also, students were encouraged to attempt the S104 quiz as they completed S154, to verify that they were now ready to move on.

Evaluation methodology

Much research into the effectiveness of e-assessment is limited to surveys of student opinion and self-reported behaviours. The large number of students studying OU level 1 science modules has presented many opportunities to seek student opinion. However there have also been opportunities to observe students actually attempting some of the questions. In addition, the vast quantity of data captured within the OpenMark system has enabled an extensive anonymous investigation into student behaviour as recorded in the data. Finally, interesting findings from the user observation and data analysis were followed-up by a further questionnaire and telephone interviews.

The sources of the data reported here were:

1. Feedback questions at the end of developmental iCMAs (offered to students on S103, the predecessor to S104, as a purely-formative add-on to the course) and at the end of 'Are you ready for level 1 science?'
2. Questionnaires sent electronically to 400 students on each of the first two presentations of S104, starting in February and October 2008, and returned by 87 students (21%) and 64 students (16%) respectively.
3. Analysis of video-recorded observations of six student volunteers completing an iCMA in the OU's Institute for Educational Technology Usability Laboratory. These students, all of whom were studying S103 at the time of the observations in 2007, are identified by the letters A to F. C and D were female; the other students were male. A verbal think-aloud protocol was used, whereby the participants were asked to talk about what they were doing and thinking, and after the evaluation session itself, each participant was asked to comment retrospectively on the reasons for their actions.
4. Anonymous analysis of data captured by the OpenMark system for iCMAs in different modules and in different types of use (summative, formative-only, diagnostic etc.). This has included an analysis of when students attempt iCMA questions, how many students attempt each question, the extent to which responses are altered in response to feedback, and the length of short-answer free-text responses. In addition, responses to individual questions have been inspected as a means of identifying possible improvements to the answer matching, feedback and question wording, and increased insight into student misconceptions.
5. Questionnaires sent electronically to 400 students on SXR103 and returned by 129 (32%). The questions were devised partly to investigate some of the unexpected findings from the usability laboratory observations and the data analysis. The decision

to survey SXR103 students was based on the fact that many of these students have also studied other modules with iCMAs used in different ways. Eight students who made interesting points in their responses to the questionnaires were interviewed in semi-structured telephone interviews in late 2009. These students, who had all studied SXR103 and other modules, are identified by the letters M to T. Q and R were female, the other students were male.

Evaluation findings

Student opinion

Tables 1 and 2 indicate the extent to which students agree or disagree with a number of statements about iCMAs. In general, it can be seen that students regard iCMAs as useful in helping them to learn and in highlighting what they need to study further, with 87% of respondents agreeing with Statement 1.1, 78% agreeing with Statement 2.1 and 79% agreeing with Statement 2.2. 64-68% agree that answering iCMA questions is fun (statements 1.5 and 2.8). These findings are substantiated by free-text survey comments from students such as

- *I think iCMAs are a great tool to help consolidate learning.*
- *it's more like having an online tutorial than doing a test.*
- *should be spread to all courses in addition to TMAs.*

Interviewed student M felt the iCMAs were particularly useful because he was studying at home and without easy access to a tutor. Although most students felt that TMAs were more useful in their learning (statements 2.3. and 2.4), a large percentage were neutral in their response to these statements and some felt that iCMAs were more useful. When this point was followed in the interviews, most people identified iCMAs and TMAs as useful for different things. R said that she would be happy with modules assessed entirely by iCMA. Student D (usability laboratory) highlighted the importance of the timeliness of iCMA feedback to

> *give you confidence that you're heading in the right lines, which you don't get at the moment. Obviously you get your TMA, but by the time you've sent it in, and got it back, it might be three weeks.*

The instantaneous receipt of feedback was the most commonly identified useful feature of iCMAs, with one student contrasting iCMAs with computer-marked assignments (CMAs) in earlier modules, which were submitted and returned through the post, and thus

> *the answers when they did return after being marked were of little interest as the course had moved on so far by that stage - the iCMA system I think is great - knowing instantly where you are going wrong.*

Other features of iCMAs that were identified as particularly useful included the availability of three attempts, the content of the feedback prompts and the references to course materials. T was pleased that the questions were relatively testing:

> *They're not walkovers, not like an American kind of multiple choice where you just go in and you have a vague idea but you know from the context which one is right.*

However it should not be forgotten that a small number of students do not find iCMAs helpful (statements 1.1, 2.1 and 2.2) or enjoyable (statements 1.5 and 2.8), perhaps linked to the fact that some (rightly or wrongly) believe that the computer sometimes marks their answers inaccurately (Statement 2.6) or penalises them for careless mistakes (Statement 2.7). A decision was taken to interview some students whose survey responses had indicated some disquiet with iCMAs. Q had said:

I discovered, through finding an error in the question, that not everybody was given the same questions. I thought this was really unfair especially as they failed to mention it at any point throughout the course.

S had said:

I have found on a number of occasions that getting an iCMA question completely wrong unholy unfair.

However, significantly, it transpired during the interviews that both of these students were extremely happy with iCMAs in general, just not with particular questions or with aspects of their use.

Most students felt that their mark for iCMA questions should count towards their overall course score (71% disagreed with Statement 2.10), and those interviewed felt that the 25% weighting in S104 was about right. It proved impossible to interview any students who felt that the mark they got in iCMAs should not count towards their overall score. There was nevertheless a difference in the reported influence of marks on behaviour, with two extremes being M, who reported engaging in summative and formative-only iCMAs in exactly the same way and T (who had not attempted SXR103's purely formative-iCMA) who said:

If the score doesn't count why on earth would you bother at all? It depends somewhat on your style of study and how much time you've got available. Now a lot of people nowadays are holding down full-time jobs and other heavy commitments. If the score doesn't count you really don't have time to mess about.

When are iCMAs attempted and how many questions are attempted?

Figures 2-5 illustrate analyses using data captured by the OpenMark system. Figures 2(b) and 2(c) illustrate the influence of cut-off date on the overall activity for summative iCMAs; use builds as the cut-off date approaches. The purely-formative practice iCMA is used throughout the entire 10-week presentation (Figure 2(a)). Formative-only iCMAs are used as practice for summative assessment, in particular end-of module examinations (Figure 3).

Figures 4(a) and 4(b) illustrate typical behaviours for individual students on all summative iCMAs. Many students behave like Student 1; they open the iCMA and attempt all 10 questions on a single day, frequently close to the cut-off date (Figure 4(a)). The behaviour shown in Figure 4(b) (Student 2) is similarly common for all summative iCMAs – students open the iCMA and look at all the questions, then they attempt them as and when they are able to as they work through the course material, using feedback from unsuccessful attempts at the iCMA questions. The behaviour shown in Figure 4(c) (Student 3) is quite typical for S154 but is not seen for other modules. Students are advised to attempt questions 1-4 after completing Chapter 2, questions 5-6 after Chapter 3 and questions 7-10 after Chapter 4, and many do precisely this.

Not surprisingly, when iCMAs are used summatively, most students complete all the questions. In formative-only use, there is typically a reduction in use as the iCMA progresses, as shown in Figure 5. As the iCMA progresses, there is both a decrease in the number of people who have completed each question (dark lines) and a decrease in the extent to which users repeat questions (paler lines). It has been suggested that this decline in use is caused by having too many questions, but a similar decline is seen for modules with several shorter formative-only iCMAs; use decreases both during each iCMA and from iCMA to iCMA. Furthermore, there are always some users who access iCMAs, but do not complete any questions. The iCMA shown in Figure 5 (where 640 users completed Question 1 and 668 completed Question 2), was opened by 768 registered students, and it appears that around 100 of these people did not take further action. In interviews, most students were happy to admit to a lower level of engagement with iCMAs when in formative-only use, but none admitted to failing to complete any questions after opening an iCMA, so the reason for this behaviour is not known (though it is speculated to occur when students decide for themselves that the questions are either trivial or too difficult). The general question-by-question reduction in use is bucked in several places, e.g. at Question 43 and Question 46. These questions are identified as the first questions in new chapters, presumably chapters that students find challenging and so seek additional practice and reinforcement. Thus, as in the summative iCMAs, clear signposting appears to be beneficial.

Use of feedback

When asked in end-of-iCMA feedback questions whether they found the feedback useful, the number who respond in the affirmative is consistently around 90%. Similarly, 85% of surveyed S104 students agreed with the statement 'If I get the answer to an iCMA question wrong, the computer-generated feedback is helpful' (Table 1 Statement 1.3) and 85% of surveyed SXR103 students agreed with the statement 'If I initially get an answer to an iCMA question wrong, the hints enable me to get the correct answer at a later attempt' (Table 2 Statement 2.5).

Observations of students in the usability laboratory painted a rather different picture. Whilst students sometimes made good use of the various aspects of the feedback provided in altering their answer for a subsequent attempt (making use of the simple fact of being told they were wrong, the more detailed prompt and the reference to the course material), there were also several instances where students did not pay sufficient attention to the feedback even when they appeared to read it. For example, student C entered 'absorbing a photon' in answer to a question for which a correct answer would have referred to the emission not absorption of a photon. Her answer was incorrect, but a software problem (later resolved) meant that she was told her answer was correct. She looked at the final answer provided by the iCMA question and said 'oh, did that right' despite the fact that the word 'emission' was clearly visible, emboldened, on the screen she was looking at.

In an attempt to investigate factors that might influence the use that students make of feedback, an analysis was performed into the extent to which incorrect responses are left unchanged for a second or third attempt after feedback has been provided and the extent to which the data-entry box is left blank. Both of these types of behaviour are more common for free-text questions than for selected response items and for questions that the student considers more difficult. Whether the iCMA is summative, formative-only or diagnostic is also an influential factor. So for variants of the same question (a question requiring the student to use a provided word equation to calculate density): in summative use, 21% of the third-attempt responses were identical to those given previously with 2% of them blank; in formative-only use, 46% of the third-attempt responses were identical to those given previously with 7% of them blank; in diagnostic use, 55% of the third attempt responses were identical to those given previously with 19% of them blank. Interviews identified a reluctance to spend time on a question (e.g. finding a calculator) when the mark did not count as a factor behind this result. In addition, students who have a complete lack of understanding of the question or feedback feel unable to enter an answer in the first place or to alter it. As T said:

> *I found that the hint in the second attempt was absolutely uninformative and I couldn't see where I was wrong*.

Improvements as a result of evaluation – two case studies

The case of an evolving question - Dawlish sandstone

The wording of the question shown to the top and left of Figure 6 has evolved over a number uses to the wording shown to the bottom and right of the figure. Although the authoring software has been altered, the basis of the answer matching has remained virtually unchanged throughout and has been consistently accurate, with approximately 99% agreement with the question author when checked against several thousand responses.

Initially there was no upper limit to the length of response that would be accepted. The mention of Dawlish in the question appears to have encouraged students to look at websites describing the geology of the Dawlish area. The resulting answers typically included a correct answer within an incorrect one (because they were describing the geology of the whole area, not just the rock that was described in the question) and were very long. One such answer was:

> This shows imbedded aeolian (windblown) sands and fluvial (water laid) breccias. Patterns of cross bedding in the sandstone shows where dunes were partly eroded and then overlain by others. This more angular nature of the breccias indicates that they were deposited by sheet flash floods and are multi storey indicating repeated flooding impact. The change from aeolian sands to the fluvial breccias may have resulted from an increase in rainfall associated with climate change at the end of the lower Permian.

Answers containing both correct and incorrect aspects present the largest challenge for the automatic marking of short-answer questions (Jordan & Mitchell, 2009) and long answers take a lot of processing time. Finally, and perhaps most significantly, the question author wanted students to work out the answer from the information provided, rather than looking on a website. For this reason the wording of the question itself was altered as shown.

In addition, for all short-answer free-text questions, responses were limited to no more than 20 words. Initially this was accompanied by the warning 'You should give your answer as a short phrase or sentence. Answers of more than 20 words will not be accepted'. Although this removed the previous ridiculously long responses, the average length of the responses actually increased for most questions (the initial median length of response for the question shown in Figure 6 was 10 words with a range of 0-129 words; this altered to a median of 13 words with a range of 0-20 words) largely because students appeared to be attempting to give answers of 19 or 20 words, even when 3 or 4 would have been adequate. For this reason, the wording of the warning was changed to that shown in Figure 6. It is only when students enter a response of more than 20 words that they are told that this is too long and given an extra attempt, with no loss of credit, in order to shorten their response. This change reduced the median response length to 10 words with a range of 0-20 words.

The wording on the box on which students have to click in order to submit an answer for checking was altered from ‘Check’ to ‘Enter answer’ in the light of evidence from the usability laboratory that some students did not understand the previous wording.

The final change to this question concerns the feedback received after a student has entered an incomplete answer at the first attempt, where that incorrect answer includes the word ‘desert’. The survey and interviews with students indicated some disquiet about the marking of this question, always caused by the fact that students had given an answer such as ‘in a desert’ which they (incorrectly) believed to be a completely correct answer. In cases like this, just being told that their answer was incorrect was not helpful, so targeted feedback has been added, as shown.

Inspection of the responses to the question shown in Figure 6 led to the identification a few identical but strangely worded responses:

> It's probably an eolian (desert/dune) deposit...the sorting, rounding, and pitting tells you that. The grains contain some iron oxide (rust), so they weren't reworked long enough to alter diagenetically to quartz.

This response was found posted on a homework site, and the same University procedures that follow high Turnitin® or Copycatch matches for TMA answers were applied.

Are you ready for level 1 science?

The diagnostic quiz ‘Are you ready for science study?’ was evaluated in summer 2009, after it had been in use for just over 2 years. It had been accessed by an average of 3200 people per month and extremely well received by prospective students, in confirming their choice of module, in guiding them towards a more appropriate choice or in helping them to decide between several possible starting points.

However, more detailed analysis of usage indicated that very many users were getting lost in the complexity of the three interlinked iCMAs, for example not clicking on the link that would take them from the introductory quiz to one of the others, and students reported wanting more direct information about whether they were sufficiently prepared for a particular module, rather than the detailed introduction about all the courses they might consider. For these reasons, ‘Are you ready for level 1 science?’ has been replaced by two separate diagnostic quizzes ‘Are you ready for S104?’ and ‘Are you ready for science study?’ (for the other level 1 science modules). Every effort has been made to reduce the number of links that prospective students need to follow in reaching and navigating the quizzes.

In addition, whilst the previous quizzes relied on students making decisions about their next step on the basis of the feedback on individual questions, 'Are you ready for S104?' also offers general feedback about their preparedness, by way of a 'traffic light' system (Figure 7).

Very few changes have been made to the individual questions in the diagnostic quizzes but, in line with the findings from other uses of iCMAs, occasional minor details were found to lead to student dissatisfaction that was out of all proportion to the ease with which the problem could be solved. For example, one question asked users to identify the chemical symbol for the element chlorine. Unfortunately, in the sans serif font used, it was impossible to distinguish the lower case L in the symbol Cl from an upper case i or the number one. An easy solution was to change the element under discussion.

Discussion

It has been shown, here and elsewhere, that e-assessment can be extremely effective in supporting learning. However student perception is very important (Gipps, 2005) and minor problems can easily lead to a loss of confidence. For this reason, it is very important to monitor the use of e-assessment and to make changes (to individual questions, the structure of an assignment or to the underlying e-assessment system) as and when appropriate. Frequently a very minor change in wording can lead to a considerable improvement in the performance of a question. The monitoring of responses to e-assessment questions can also provide information about common student errors, which can in turn lead to more appropriate feedback, to changes in the method of teaching and to more widely applicable insight into student misconceptions (Jordan, 2006).

Bull & Danson (2004) point out that new forms of assessment can attract criticism rarely considered in the traditional assessment process. Anxiety has been expressed that iCMAs are open to plagiarism, but this should be no more the case than any assignments which are not completed under examination conditions. Plagiarism can be encouraged by poorly designed assessment of any type and discouraged by the careful wording of questions. The use of different variants of iCMA questions can reduce opportunities for collusion, and plagiarism detection software could be applied to short-answer free-text iCMA responses.

Issues of accessibility have also been raised. Ball (2009) reminds us that the use of e-assessment should in general improve accessibility, and OpenMark and Moodle both have excellent accessibility features, including a range of font sizes and colours and a text version of most questions (which can be fed into a screen-reader). However Richardson (2010) rightly points out the dangers of adopting a 'one-size fits all' approach in introducing new forms of educational technology. A small number of students, in particular offender learners, cannot access the internet at all – the inexorable increase in the use of online learning and assessment, however beneficial for most students, runs a danger of further excluding the already socially excluded (Watts, 2010).

Much consideration has been given to the impact of the measuring (summative) function of assessment on learning, with Gibbs (2010) concluding that Open University assignments would be more effective if they gave feedback but no marks. However research into the weighting or otherwise of e-assessment specifically is less well developed. The findings that have been reported in this paper are in line with those of Kibble (2007), namely that a light summative weighting can increase student engagement with e-assessment, but that the summative function may overwhelm the formative.

Similarly, much effort has gone into identifying conditions under which students make effective use of feedback, but student interaction with feedback delivered by a computer is less well understood. Nix and Wyllie (in press) found that student engagement with feedback depends on the student's learning style, their confidence and whether their answer was right or wrong.

Conclusions and suggestions for further work

Interactive computer-marked assessment has huge potential to help distance learners to find appropriate starting points, to provide them with timely feedback and to help them to pace their study. Sophisticated question types require students to construct a response of their own rather than merely selecting from a number of pre-determined options. But how far is it appropriate to go? Automatic essay-marking software such as E-rater® (Attali & Burstein, 2006) is quite commonly used in the USA, but regarded with some caution elsewhere. For the foreseeable future, e-assessment is best regarded as just one of a number of assessment and teaching tools, with tutor-marked assignments remaining the appropriate vehicle for assessing higher-order learning outcomes and communication skills. The use of e-assessment does not imply a tutor-less future, rather one in which tutors are freed from the drudgery of marking simple items to give increased support for their students, with information about student performance on e-assessment tasks used to encourage dialogue.

Further work is needed into factors that affect student engagement with e-assessment and their use of computer-generated feedback. The vast quantity of data collected from online systems enables rigorous quantitative analysis of student interaction with e-assessment, going far beyond the all too common surveys of student opinion and self-reported behaviour.

Acknowledgements

The author gratefully acknowledges the work of Linda Fowler and Ruth Williams in developing 'Are you ready for level 1 science?', and frequent assistance from media developers Greg Black and Fiona Thomson. Without the inspiration and commitment of Phil Butcher, E-assessment Adviser, much of the work described in this paper could not have taken place.

References

Attali, Y & Burstein, J. (2006) Automated essay scoring with e-rater® V.2 *JTLA: The Journal of Technology, Learning & Assessment*, 4(3), 1-30.

Ball, S. (2009) Accessibility in e-assessment, *Assessment & Evaluation in Higher Education*, 34(3), 293-303.

Black, P. & Wiliam, D. (1998) Assessment and classroom learning, *Assessment in Education*, 5(1), 7-74.
Boud, D. (1995) Assessment and learning: contradictory or complementary?, in: P. Knight (Ed) *Assessment for learning in higher education* (London, Kogan Page), 35-48.
Boud, D. (2000) Sustainable assessment: rethinking assessment for the learning society, *Studies in Continuing Education*, 22(2), 151-167.
Bull, J. & Danson, M. (2004) *Computer-aided assessment (CAA)* (York, LTSN Generic Centre).
Butcher, P.G. (2008) Online assessment at the Open University using open source software: Moodle, OpenMark and more. 12th International CAA Conference, Loughborough, UK.
Butcher, P.G & Jordan, S.E. (2010) A comparison of human and computer marking of short free-text student responses. *Computers & Education*, 55, 489-499.
Dysthe, O. (2008) The challenges of assessment in a new learning culture, in: A. Havnes & L. McDowell (Eds) *Balancing dilemmas in assessment and learning in contemporary education* (London, Routledge), 15-28.
Freake, S. (2008) Electronic marking of physics assignments using a tablet PC, *New Directions*, 4, 12-16.
Gibbs, G. (2010) Does assessment in open learning support students?, *Open Learning*, 25(2), 163-166.
Gibbs, G. & Simpson, C. (2004) Conditions under which assessment supports students' learning, *Learning and Teaching in Higher Education*, 1, 3-31.
Gipps, C.V. (2005) What is the role for ICT-based assessment in universities?, *Studies in Higher Education*, 30(20), 171-180.
Hattie, J. & Timperley, H. (2007) The power of feedback, *Review of Educational Research*, 77, 81-112.
Havnes, A. & McDowell, L. (Eds) (2008) *Balancing dilemmas in assessment and learning in contemporary education* (London, Routledge).
Jordan, S. (2007) The mathematical misconceptions of adult distance-learning science students. Proceedings of the CETL-MSOR Conference, Loughborough, 11-12 September 2006. Maths, Stats and OR Network, 87-92.
Jordan, S. & Mitchell, T. (2009) E-assessment for learning? The potential of short free-text questions with tailored feedback, *British Journal of Educational Technology*, 40(2), 371-385.
Kibble, J. (2007) Use of unsupervised online quizzes as formative assessment in a medical physiology course: effects of incentives on student participation and performance. *Advances in Physiology Education*, 31, 3, 253-260.
Kirkwood, A. & Price, L. (2008) Assessment and student learning: a fundamental relationship and the role of information and communication technologies, *Open Learning*, 23 (1), 5-16.
Kornell, N. , Jenson Hays, M. & Bjork, R.A. (2009) Unsuccessful retrieval attempts enhance subsequent learning, *Journal of Experimental Psychology*, 35(4), 989-998.
Mackenzie, D. (2003) *Assessment for e-learning: what are the features of an ideal e-assessment system?* 7th International CAA Conference, Loughborough.
Mackenzie, D. (2004) *Online assessment: quality production and delivery for higher education.* Available online at

www.enhancementthemes.ac.uk/documents/events/20040416/Mackenziepaper-revised.pdf (accessed 2nd July 2010).
Nicol, D.J. (2007) E-assessment by design: using multiple choice tests to good effect, *Journal of Further and Higher Education*, 31(1), 53-64.
Nicol, D.J. & Macfarlane-Dick, D. (2006) Formative assessment and self-regulated learning: a model and seven principles of good feedback practice, *Studies in Higher Education*, 31(2), 199-218.
Nix, I. & Wyllie, A. (in press) Exploring design features to enhance computer-based assessment: Learners' views on using a confidence-indicator tool and computer-based feedback, *British Journal of Educational Technology*.
Richardson, J.T.E. (2010) Course completion and attainment in disabled students taking courses with the Open University UK, *Open Learning*, 25(2), 81-94.
Ross, S.M., Jordan, S.E., and Butcher, P.G. (2006) Online instantaneous and targeted feedback for remote learners, in: C. Bryan and K.V.Clegg (eds) *Innovative assessment in higher education* (London, Routledge), 123-131.
Scouller, K.M. & Prosser, M. (1994) Students' experiences of studying for multiple choice question examinations, *Studies in Higher Education*, 19(3), 267-279.
Sim, G., Holifield, P. & Brown, M. (2004) Implementation of computer assisted: lessons from the literature, *ALT-J*, 12(3), 215-229.
Snyder, B.R. (1971) *The hidden curriculum* (New York, Alfred A. Knopf).
Watts, J.H. (2010) Teaching a distance higher education curriculum behind bars, *Open Learning*, 25(1), 57-64.
Whitelock, D. & Brasher, A. (2006). Roadmap for e-assessment. Available online at www.jisc.ac.uk/elp_assessment.html (accessed 2nd July 2010)

Table 1 Reaction to statements about iCMAs from 87 students on the first presentation of S104 and 64 students on the second presentation (data combined for all 151 students).

Statement	**DA**	**MA**	**N**	**MD**	**DD**
1.1 Answering iCMA questions helps me to learn.	80 (54%)	49 (33%)	7 (5%)	8 (5%)	4 (3%)
1.2 The mark I am awarded for each iCMA helps me to gauge how well I am doing.	52 (35%)	55 (37%)	22 (15%)	14 (9%)	5 (3%)
1.3 If I get the answer to an iCMA question wrong, the computer generated feedback is helpful.	71 (47%)	57 (38%)	11 (7%)	8 (5%)	3 (2%)
1.4 I think the computer sometimes marks my iCMA answers wrong.	16 (11%)	18 (12%)	34 (23%)	36 (24%)	46 (31%)
1.5 Answering iCMA questions is fun.	44 (30%)	51 (34%)	34 (23%)	7 (5%)	13 (9%)

Note: DA is definitely agree, MA is mostly agree, N is neither agree nor disagree, MD is mostly disagree, DD is definitely disagree. Percentages are of the number of respondents to each question and do not always add up to 100% because of rounding.

Table 2 Reaction to statements about iCMAs from 129 students on SXR103, all but 15 of whom had also used iCMAs on other modules.

Statement	**SA**	**MA**	**N**	**MD**	**SD**	**N/A**
2.1 Answering iCMA questions directly helps me to learn skills or knowledge.	48 (39%)	49 (40%)	11 (9%)	8 (6%)	4 (6%)	4 (6%)
2.2 Answering iCMA questions helps me to understand what I need to study further.	48 (38%)	51 (41%)	12 (10%)	5 (4%)	4 (3%)	5 (4%)
2.3 I learn more by doing TMA questions than by doing iCMA questions.	22 (18%)	32 (26%)	43 (34%)	22 (18%)	3 (2%)	3 (2%)
2.4 The feedback that my tutor provides on my TMA answers is more helpful than the feedback provided by the computer on my iCMA answers.	27 (22%)	46 (37%)	33 (26%)	14 (11%)	1 (1%)	4 (3%)
2.5 If I initially get an answer to an iCMA question wrong, the hints enable me to get the correct answer at a later attempt.	29 (23%)	78 (62%)	7 (6%)	8 (6%)	2 (2%)	2 (2%)
2.6 The computer sometimes marks my iCMA answers incorrectly.	4 (3%)	18 (14%)	21 (17%)	23 (18%)	48 (38%)	11 (9%)
2.7 The computer unfairly penalises me for careless mistakes.	11 (9%)	16 (13%)	27 (21%)	38 (30%)	28 (22%)	6 (5%)
2.8 Answering iCMA questions is fun.	18 (14%)	68 (54%)	27 (21%)	10 (8%)	2 (2%)	2 (2%)
2.9 My performance in iCMA questions reflects my ability.	18 (14%)	63 (50%)	28 (22%)	15 (12%)	2 (2%)	1 (1%)
2.10 I don't think that the mark I get in iCMAs should count towards my overall course score.	5 (4%)	9 (7%)	21 (17%)	33 (26%)	57 (45%)	1 (1%)

Note: SA is strongly agree, MA is mostly agree, N is neither agree nor disagree, MD is mostly disagree, SD is strongly disagree. Percentages are of the number of respondents to each question and do not always add up to 100% because of rounding.

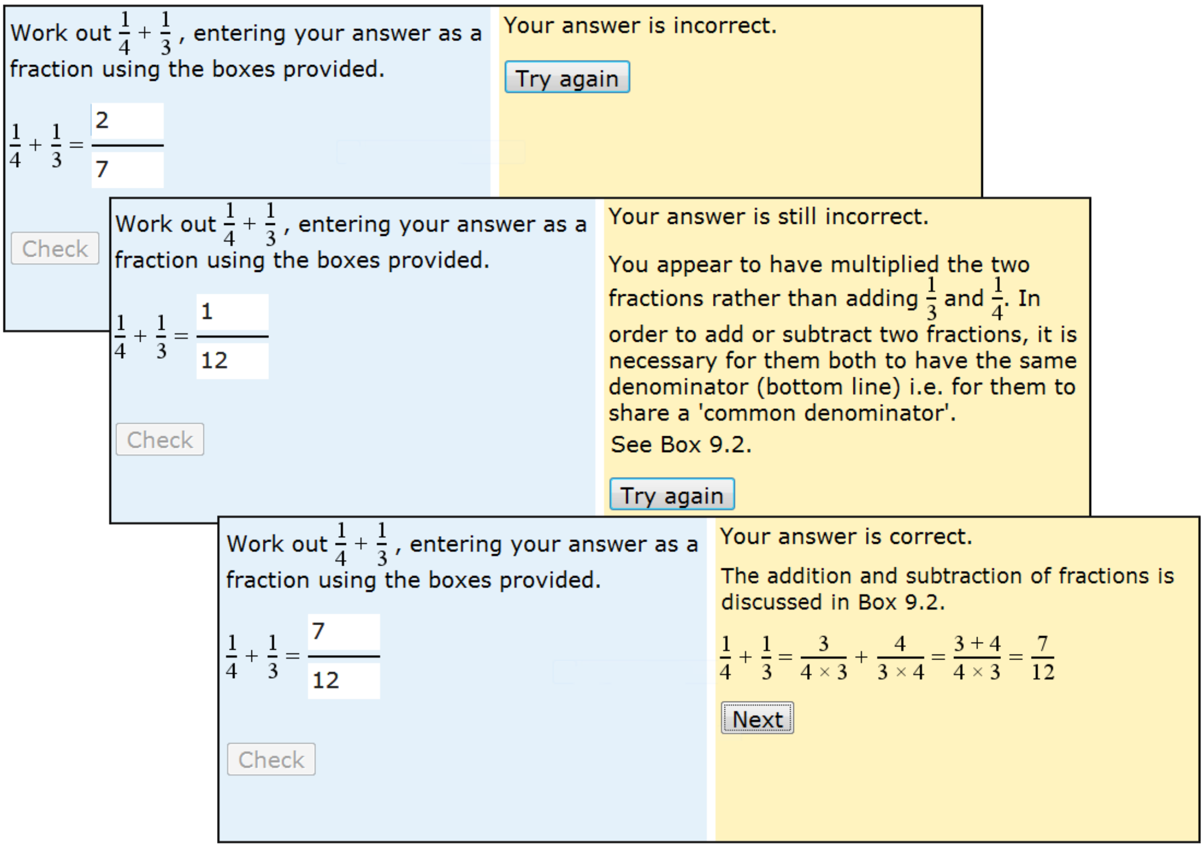

Figure 1 An OpenMark interactive computer-marked assignment question, showing increasing feedback after repeated attempts.

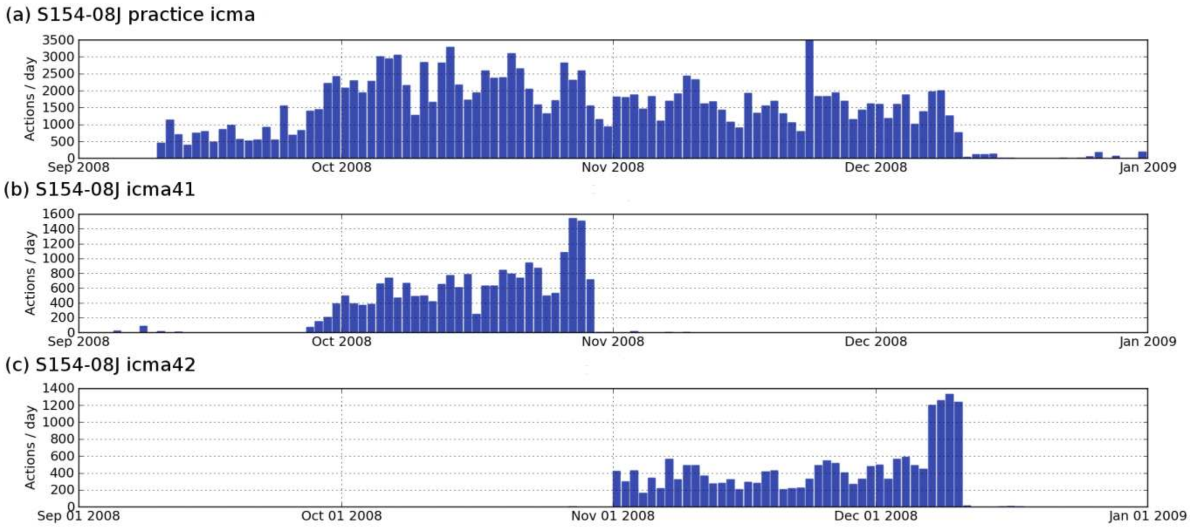

Figure 2 Action levels (total number of interactions for all students) for (a) the S154 practice iCMA for the presentation that started on 27th September 2008 and (b) and (c) the two summative iCMAs (iCMA41 and iCMA42) for the same presentation. The cut-off date for iCMA41 was 29th October 2008; the cut off date for iCMA42 was 10th December 2008.

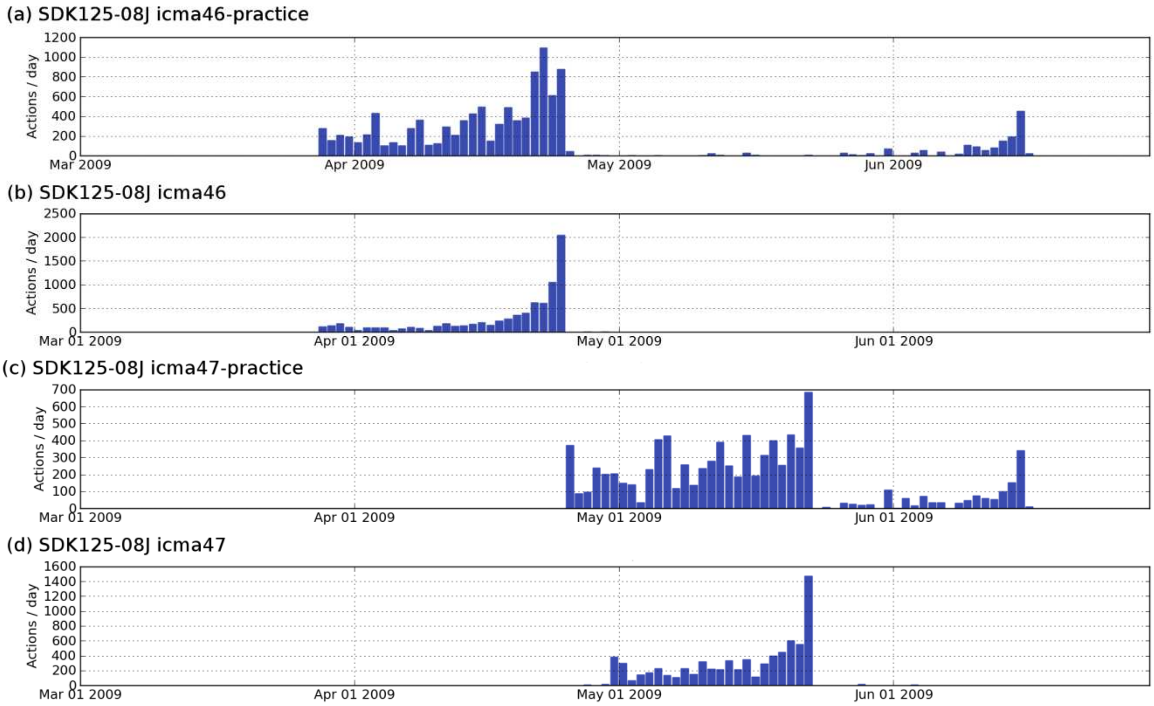

Figure 3 Action levels (total number of interactions for all students) for (a) and (c) two of SDK125's practice iCMAs; (b) and (d) the equivalent summative iCMAs. The cut-off date for iCMA46 was 24th April 2009 and the cut-off date for iCMA47 was 22nd May 2009; the examination was on 16th June 2009.

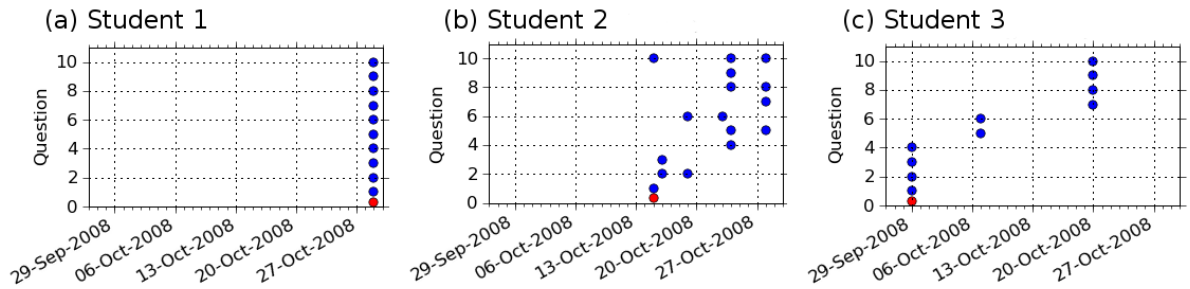

Figure 4 Three typical student behaviours exhibited on S154 summative iCMAs. The cut-off date for this iCMA was 29th October 2008.

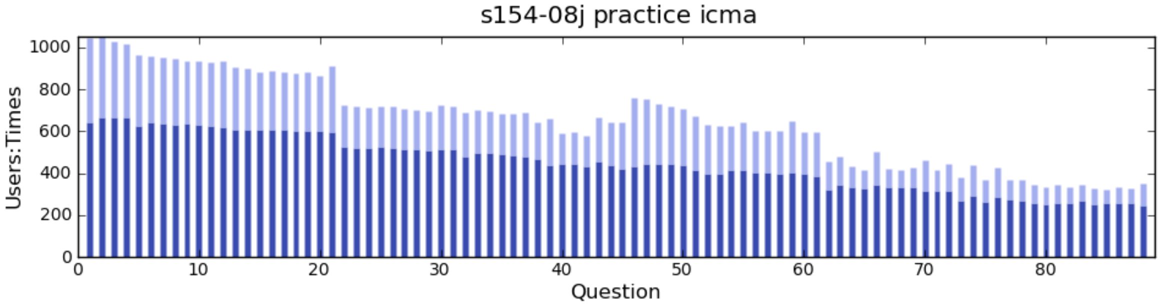

Figure 5 The number of students who completed each question in an S154 practice iCMA. The dark lines show the number of individual students who completed each question; the paler lines show the total number of completions for each question, thus indicating the extent to which questions were repeated.

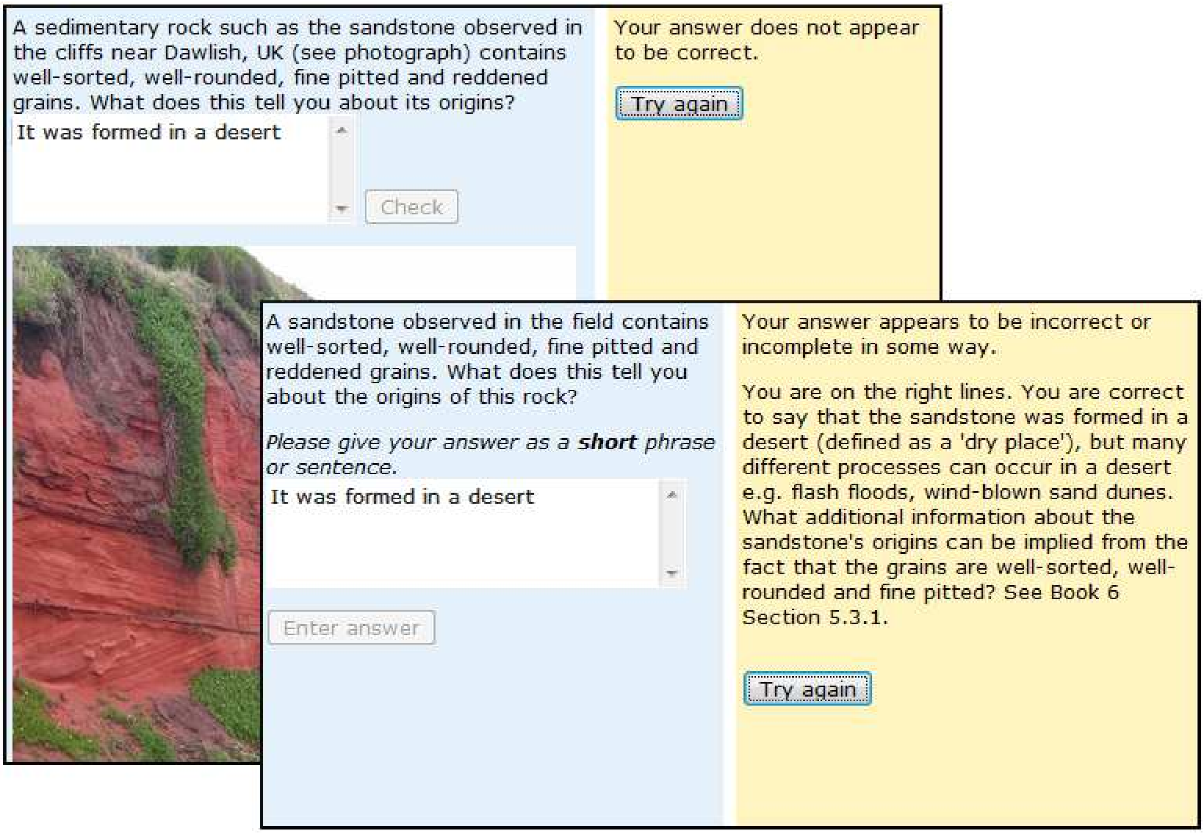

Figure 6 The original and current versions of an S104 iCMA question.

The Open University

Are you ready for S104?

Results for A12345678

Display options | Help

This feedback is designed to give you some indication, based on your answers to *Are you ready for S104?* of your preparedness to study Exploring science (S104). The feedback is based on a 'traffic lights' system.

If you receive all greens then you appear to have sufficient time and previous knowledge to study S104.

If you receive one or more ambers then you may need to find more time or to revise your skills in one or more areas before taking S104. You may wish to consider studying *Science starts here* (S154) before studying S104. S154 will prepare you fully for studying S104 in the future.

If you receive one or more reds then you do not appear to be sufficiently prepared to study S104 at the present time. You are advised to study S154 *Science starts here* before studying S104. S154 will prepare you fully for studying S104 in the future.

You should also note that, however well prepared you are, you will need to find about 16 hours a week over a 9-month period in order to study S104 successfully.

If you would like more information about S154, please visit the *Science starts here* (S154) webpage.

If you would like to check whether you are sufficiently prepared to study S154, please attempt the *Are you ready for science study?* questions.

Essential English for S104
You achieved : Amber

Essential Mathematics for S104
You achieved : Green

Valuable Science for S104
You achieved : Amber

Valuable Mathematics for S104
You achieved : Amber

Figure 7 The general feedback provided to users of 'Are you ready for S104?' on the basis of their answers to the diagnostic questions.