Using e-assessment to support distance learners of science

Conference or Workshop Item

Using e-assessment to support distance learners of science

Sally Jordan and Philip Butcher

OpenCETL, The Open University, Milton Keynes, United Kingdom, MK7 6AA

Contact email: s.e.jordan@open.ac.uk

Introduction

This paper describes the use of online interactive computer marked assignments (iCMAs) in a range of physics and general science modules at the UK Open University (OU), and an evaluation, jointly funded by the Physics Innovations Centre for Excellence in Teaching and Learning (piCETL) and the Centre for Open Learning of Mathematics, Science, Computing and Technology (COLMSCT) into student engagement with these iCMAs. Although the work has been done in the context of one particular large distance-learning university, the software used is open source and the evaluation has wider implications for those seeking to find the most appropriate way to use e-assessment to support their students' learning.

The Joint Information Systems Committee (JISC) define e-assessment as 'the end-to-end electronic assessment processes where ICT is used for the presentation of assessment and the recording of responses'(JISC, 2009). There are many exciting ways in which e-assessment can be used to support learning, including the use of e-portfolios and the assessment of students' contributions to online discussions (Buzzetto-More and Alade, 2006). Here we describe the use of online quizzes, where a student enters his or her responses and receives instantaneous feedback. Note however that the range of question types in use goes far beyond the selected response multiple-choice questions that are commonly associated with assessment of this type.

Background

Throughout the Open University's 40-year history there has been some blurring of the summative ('for measuring') and formative ('for learning') roles of assessment. OU undergraduate students are typically (though no longer exclusively) adults, studying part-time alongside other commitments and they have a wide range of entry qualifications from previous higher education qualifications to, literally, none. Many have not studied for many years and so they may be particularly lacking in confidence. The students study at a distance, but the OU's model of supported distance education means that they are usually supported by a tutor. This tutor will provide occasional tutorials (face-to-face or using a range of synchronous and asynchronous electronic communication technologies) and be available to support student learning by telephone or email; however a substantial part of the tutor's contracted time will be spent in grading and providing feedback on 'tutor-marked assignments' (TMAs). The fact that this task is described as 'correspondence tuition' reflects the importance that is placed on the feedback provided by tutors to their students in this way; this is particularly important in a distance-learning organisation, where many students never meet their tutor and opportunities for informal

feedback are extremely limited. However TMA scores have usually contributed substantially to students' overall course grades.

The use of e-assessment also has a long history at the Open University. TMAs have long been supplemented by computer-marked assignments (CMAs), initially comprising solely multiple choice questions, with students' responses entered on machine-readable forms and submitted by post. Now, in many Science Faculty courses, students complete online interactive computer-marked assignments (iCMAs) from their own computers at home.

Why e-assessment, or why not?

It is widely recognised that rapidly received feedback on assessment tasks has an important part to play in underpinning student learning, encouraging engagement and promoting retention (see, for example, Rust, Donovan and Price, 2005). Online assessment, with its instantaneous feedback, can been seen as providing 'a tutor at the student's elbow' (Ross, Jordan and Butcher, 2006, p125), of particular importance in a distance-learning environment. For high-population modules and programmes, e-assessment can also deliver savings of cost and effort. Finally, e-assessment is the natural partner to the growth industry of e-learning.

However, opinions of e-assessment are mixed and evidence for its effectiveness is inconclusive; indeed e-assessment is sometimes perceived as having a negative effect on learning (Gibbs, 2006). There are more widely voiced concerns that e-assessment tasks (predominantly but not exclusively multiple-choice) can encourage memorisation and factual recall and lead to surface-learning, far removed from the tasks that will be required of the learners in the real world (Scouller and Prosser, 1994).

Thus when e-assessment is used, careful evaluation is required to ensure that it is being used to optimum effect and having a positive not a detrimental effect on student learning. There are a number of related questions, for example:

- What sorts of e-assessment tasks have the best potential to support student learning?
- What mode of use is most effective: summative, formative, thresholded etc.
- What sort of feedback on e-assessment tasks is most useful?
- Does it matter that the feedback is generated by a computer rather than by tutors, peers or the students themselves?

E-assessment at the Open University

The iCMAs included in the current evaluation all use the 'OpenMark' web-based assessment system (Butcher, 2006) within the Moodle virtual learning environment (Butcher, 2008). Question types include those requiring free text entry of numbers, letters, and words in addition to more conventional drag and drop, multiple choice, multiple response and hotspot questions. In most cases, students are allowed multiple attempts (usually three) at each question, with increasingly detailed and tailored prompts allowing them to act on the feedback whilst it is still fresh in their minds and so to learn from it (Sadler,1989; Gibbs and Simpson, 2004), as illustrated in the simple question

shown in Figure 1. The hints frequently refer students back to relevant course material (which might be a book, a video sequence or an interactive activity). Feedback can also be provided on the student's demonstration of learning outcomes developed in the preceding period of study. OpenMark has good accessibility features and wherever possible, questions exist in several variants. In summative use this enables different students to receive different assignments, whilst in formative-only use, the different questions provide extra opportunities to practise.

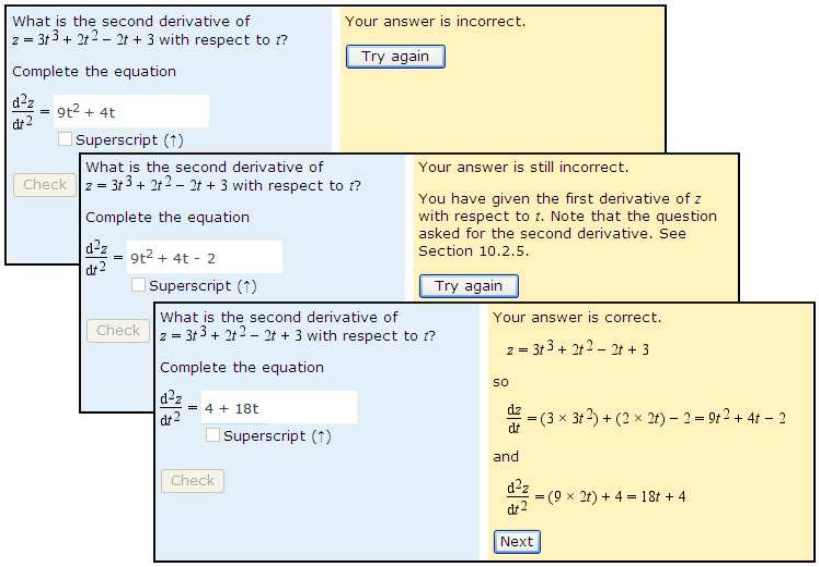

Figure 1 A simple numerical OpenMark question, with increasing feedback given at each attempt.

The range of question types has been extended still further to include those in which students have to construct a response in the form of a phrase or sentence of up to 20 words. The answer matching for these questions was initially written using a commercial linguistically-based authoring tool (Jordan and Mitchell, 2009); answer matching of equivalent accuracy can now be obtained using OpenMark's own 'PMatch' algorithmically-based system. Whichever software is used, the answer matching is developed using previous responses from students, and the questions sit within the OpenMark framework, so multiple attempts are allowed, with increasing feedback (Figure 2).

Within the Open University Science Faculty, iCMAs are embedded in courses' assessment strategies in a wide variety of ways, for example:
Case 1: The level 1 10 CATs point course S151: *Maths for Science* has a single summative OpenMark end of course assignment (available to students for 5 weeks) with instantaneous feedback given to students on individual questions. The students are not told their mark. A similar practice assessment (Case 1 PA) is available for the duration of the course.
Case 2: The introductory 10 CATs point course S154 : *Science Starts Here* has two very short tutor marked assignments and two very short summative but low stakes iCMAs

(Case 2 SA) plus a purely formative iCMA (Case 2 PA) available for the duration of the course.

Case 3: The Science Faculty's major 60 point foundation course S104 : *Exploring Science* (the introduction to physics, chemistry, biology and Earth science) has nine summative but low stakes iCMAs, eight tutor marked assignments and an end of course assignment.

Case 4: The level 3 physics course SM358 : *The quantum world* has regular formative iCMAs, clearly embedded within the course alongside tutor-marked assignments. Students are told that there will be similar questions to those in the iCMAs in the final examination. From 2010, all TMAs and iCMAs on this course and other level 3 physics and astronomy courses will be formative only, but thresholded i.e. students will have to complete a certain number of the assignments, where satisfactory completion of an iCMA will be deemed to be a score of 30% or more.

Case 5: iCMAs are used for diagnostic purposes in a series of 'Are you ready for?' quizzes, designed to help students to decide whether or not they are sufficiently prepared to study a particular course.

The Moodle Gradebook enables students to monitor their own progress, encouraging sustainable self-assessment practices (Boud, 2000), and the tutor's view of the Moodle Gradebook encourages discussion between students and tutors.

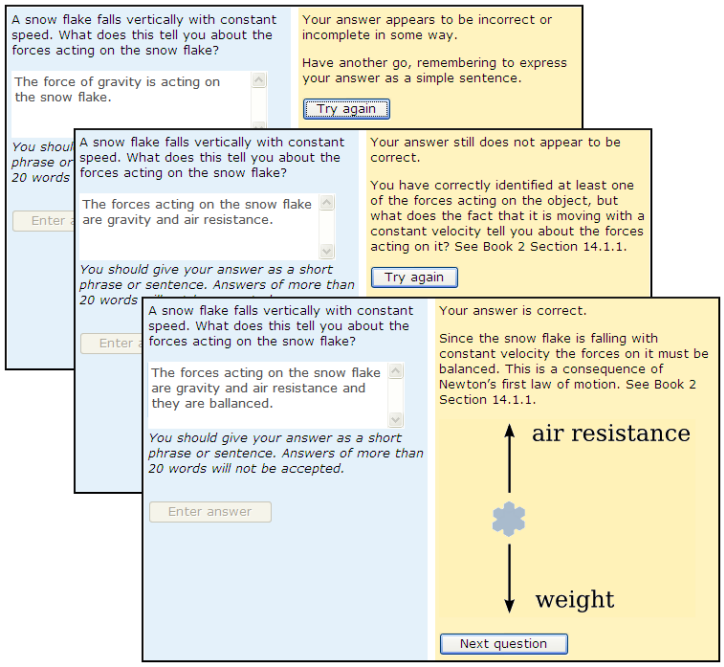

Figure 2 A short answer free-text question, with increasing feedback given at each attempt.

Previous evaluation

Evaluation methodologies have included surveys of student opinion, observation of students in a usability laboratory, a 'success case method' approach (Hunter, 2008) and a comparison of accuracy of computer and human marking (Jordan and Mitchell, 2009).

The systems have been found to be robust and accurate in marking and most students report enjoying the iCMAs and finding them useful. However there are some anomalies. For example, whilst more than 90% of students report finding the feedback provided useful and, when observed in a usability laboratory, some students were seen to read the feedback and then to adjust their answer in a sensible way, others do not make use of the feedback in the way that iCMA authors would hope.

The current work

The current project is seeking to 'observe' student behaviour remotely, by means of a quantitative and anonymised analysis of the data captured when students attempt iCMAs. Tools have been produced to extract summary information from the databases. It should be noted that the quantitative data extracted in this way are reliable since the student populations of the courses in question are large, for example, the course identified as 'Case 3' above has two presentations each year with 1500-2000 students per presentation.

How many students attempt each question?

Not surprisingly, when iCMAs are summative (even if low stakes), students are highly motivated to attempt all the questions, as shown in Figure 3 below.

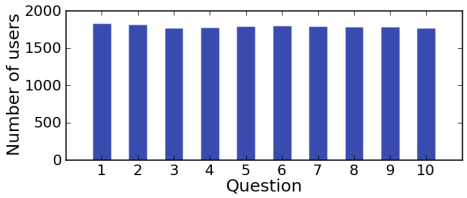

Figure 3 The number of students attempting each question in a summative but low-stakes iCMA (Case 3).

However, when an iCMA is formative-only, usage drops off in a characteristic way, as shown in Figure 4.

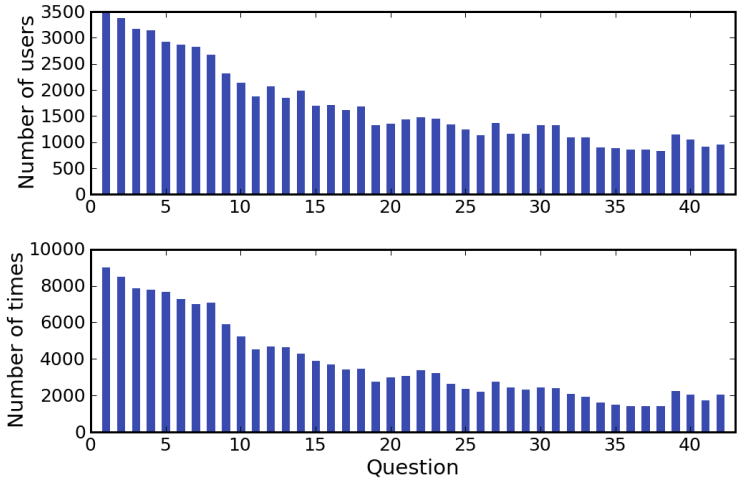

Figure 4 The number of students attempting each question in a formative-only iCMA (Case 1 PA)

The top bar-chart in Figure 4 shows the number of students who attempted each question; the lower bar-chart shows the number of separate uses of each question (so each user attempted each question an average of three times). Note that this particular iCMA includes forty-two questions; usage drops off less for iCMAs with fewer questions, however if there are then several separate iCMAs spread over the duration of the course, there is then a drop in use from iCMA to iCMA, resulting in a similar drop in use from the first question in the first iCMA to the final question in the final iCMA. Typically, the number of users has dropped to around half by half-way through the iCMA or course and to around a quarter by the end. In addition, some students view the questions in the iCMA but never attempt any; for the iCMA illustrated in Figure 4 and over the same time-scale, the iCMA was viewed by around 4500 students.

There appear to be particular aspects of iCMA design that can contribute to a marked decline in use (which is not recovered in subsequent questions); this is often linked to questions that are deemed to be complicated (perhaps with multiple boxes to complete) or time-consuming (though not necessarily difficult) or which require the student to access a piece of multi-media or even perhaps just to use their calculator (as illustrated in the example shown in Figure 5 below, which is question 19 in the iCMA under consideration in Figure 4). However use can be encouraged by making it clear which questions relate to which section of the course (as shown in the navigation panel to the left hand side of Figure 5); Figure 4 shows that students who had not attempted previous questions were nevertheless sufficiently motivated to attempt the questions linked to Chapter 7 (starting with question 27) and Chapter 10 (starting with question 39). Identifying which questions relate to which sections of the course (and reminding students to attempt them at appropriate times, by notes in the course texts or website) is now considered to be good iCMA design, although it is not practical in all situations, for instance when iCMA questions have a deliberately synoptic role.

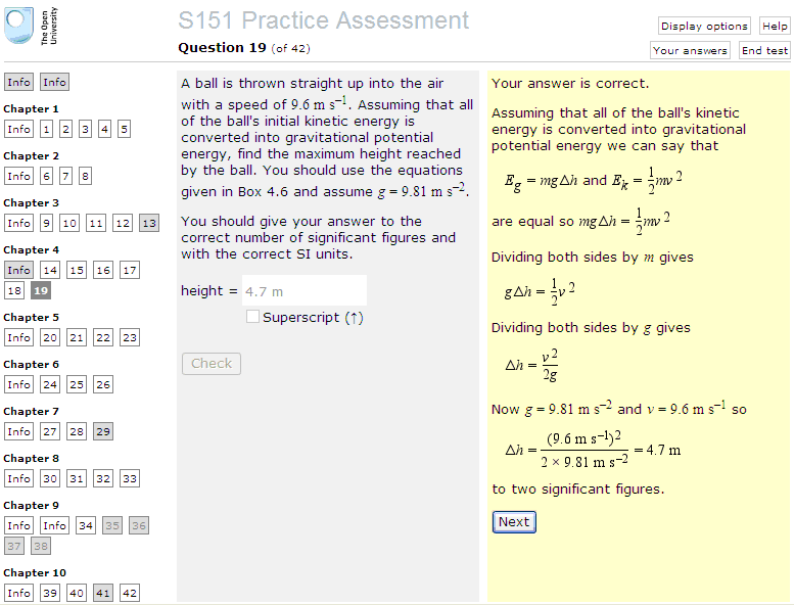

Figure 5 The iCMA under consideration in Figure 4, showing Question 19 and the navigation panel.

When do students attempt the questions?

Summative iCMAs are usually made available to students for a period of several weeks and within that time scale students are allowed to spend as long as they would like to on the questions; if a student closes their browser and returns to an iCMA at a later stage, the system will remember where they were up to. However most summative iCMAs have a 'hard' cut-off date; this is a deliberate policy, designed to encourage OU students (who frequently have many competing demands on their time) to keep up to date in their studies. In Case 3, the cut-off date for each iCMA is a few days after students are scheduled to have completed their study of the relevant course material. Figure 6 shows that the cut-off date is clearly effective in encouraging students to attempt the iCMA but most complete the questions in the few days immediately before the cut-off date. The three graphs in Figure 6 are for an early question (Question 1), a late question (Question 9) and the combined usage of all 10 questions in the iCMA; thus the behaviour is similar for all questions.

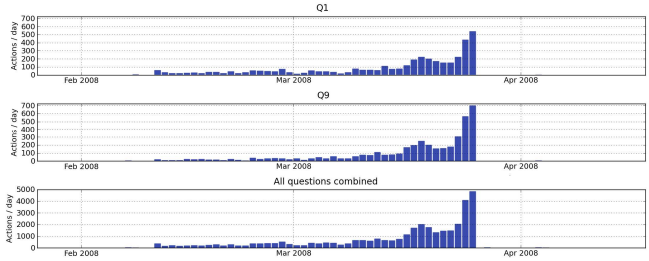

Figure 6 Number of actions on iCMA questions per day, for a summative iCMA with a hard cut-off date (Case 3)

The situation for Case 2 is rather different. The purely formative practice iCMA has 88 questions and is available to students throughout the course's 10-week duration. Figure 7 shows the number of actions per day for an early question (Question 2), a question around half-way through the iCMA (Question 40), a late question (Question 88) and all the questions combined. The relatively uniform overall usage appears to be attributable to the fact that students are attempting different questions at different times. This iCMA, like the one shown in Figure 5, has questions linked to different chapters of the course, and students are reminded after each chapter to try the relevant questions.

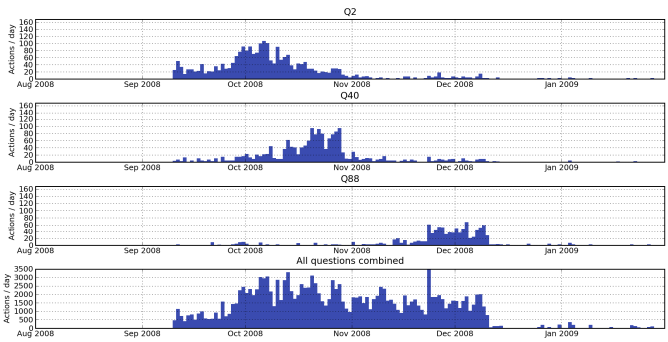

Figure 7 Number of actions on iCMA questions per day, for a formative-only iCMA (Case 2 PA)

Figure 8 shows the number of actions per day for the summative iCMA for the same course (Case 2 SA), for an early question (Question 1) a late question (Question 9) and all the questions combined. The course teams who produced the courses in Case 2 and Case 3 had designed the summative iCMAs of the two courses (which are linked; Case 2 is a precursor to Case 3) to be similar; the iCMAs have similar weightings, both have 10 questions, they both have hard cut-off dates and they are available to students for a similar length of time. So it is surprising that Figure 6 is rather different from Figure 8; in the latter case students again appear to be attempting different questions at different times. One possible explanation of this is purely that this is what students are advised to do; the questions in the first summative iCMA assess three chapters of the course; on completing Chapter 2, students are advised to attempt the relevant formative and summative questions, and similarly for Chapters 3 and 4.

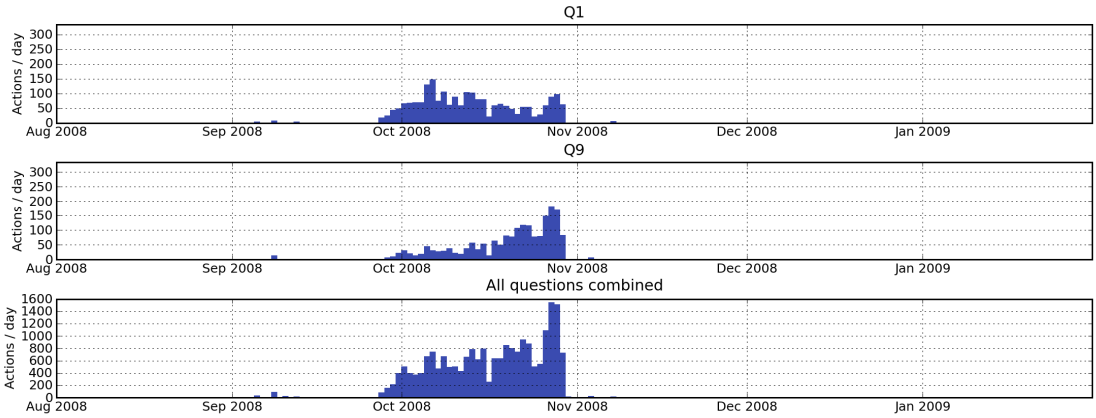

Figure 8 Number of actions on iCMA questions per day, for a summative iCMA with a hard cut-off date (Case 2 SA)

Most of the courses included in the study have a final assessed component that is completed in the student's own home and with access to course material. Case 4 (the level 3 physics course) ends with an examination, taken at an exam centre. It is clear from Figure 9 that many students are making use of the iCMA questions on this course for revision, so it appears that use of formative iCMAs can be encouraged simply by saying that the practice and feedback will be useful for the examination. Figure 9 shows the use of an iCMA primarily intended for use earlier in the year; selections of earlier questions had been combined together into three 'revision' iCMAs and these were also heavily used in the run up to the examination, though the fact that some students chose to use the earlier iCMAs perhaps indicates that they were using iCMAs in their revision of specific topics. The reasons for this behaviour will be explored further by interview.

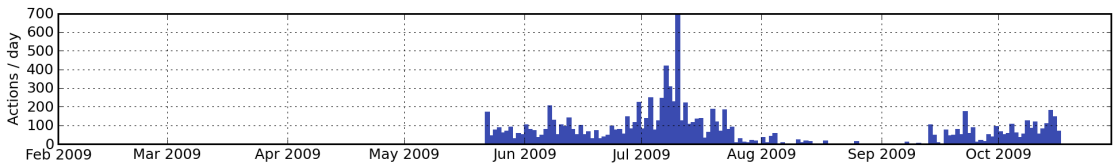

Figure 9 Number of actions on iCMA questions per day, for a formative-only iCMA on a course with a final exam (Case 4)

When do individual students attempt the iCMA?
In addition to looking at all actions on a particular iCMA question, it is possible to inspect the way in which individual students progress through the iCMA, and three typical student behaviours are shown in Figure 10 (all for Case 2 SA). 'Student 1' and 'Student 2' are typical for all summative uses: many students attempt all 10 questions on the very last day an iCMA is open (like Student 1) whilst some attempt a question, then attempt another, then return to the first question etc. in an apparently chaotic fashion, sometimes with a period of several days between consecutive uses of the same question (like Student 2). However graphs such that shown for Student 3 were observed frequently for Case 2 SA but never for Case 3, and this is another illustration of the behaviour shown in Figure 8. This student has attempted the 4 questions that assess Chapter 2, then presumably worked through Chapter 3 and attempted that chapter's questions, then similarly for Chapter 4.

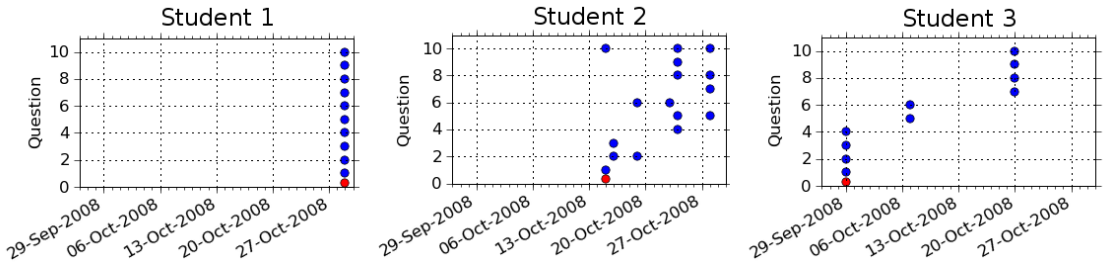

Figure 10 Days on which three students made attempts at the questions on a summative iCMA (Case 2 SA)

Closer inspection of student responses to questions
Inspection of the actual responses entered by students, in particular to free-text responses in summative use, has been used to learn about common student misconceptions (Jordan, 2007). Student responses can also provide valuable insights into more general factors concerning the use of iCMAs by students.

Closer inspection of student responses to questions: (a) the length of free-text answers
Student responses to short-answer free text questions in summative use have generally been found to be

- more likely to be correct
- more likely to be expressed as sentences (as requested in the course guide and the introductory screen in Case 3)
- longer

than the responses to the questions in formative-only use. Figures 11 and 12 compare the length of responses obtained to the question shown in Figure 2 (without the wording shown in italics).

In formative-only use (Figure 11) the peak at one word corresponds to the answer 'balanced' and the peak at 3 words corresponds to the answers such as 'they are balanced' and 'equal and opposite'. In summative use (Figure 12) the peak at 4 words

corresponds to answers such as ‘the forces are balanced’ and the peak at 8 words corresponds to answers such as ‘the forces acting on the snowflake are balanced’ and ‘there are no unbalanced forces acting on it’. It is quite common for the distribution of lengths to be bimodal; in other questions there is sometimes a peak for answers that simply answer the question (e.g. ‘the force is reduced by a factor of four’) and another for answers that add an explanation (e.g. ‘the force is reduced by a factor of four since it depends on dividing by the square of the separation’).

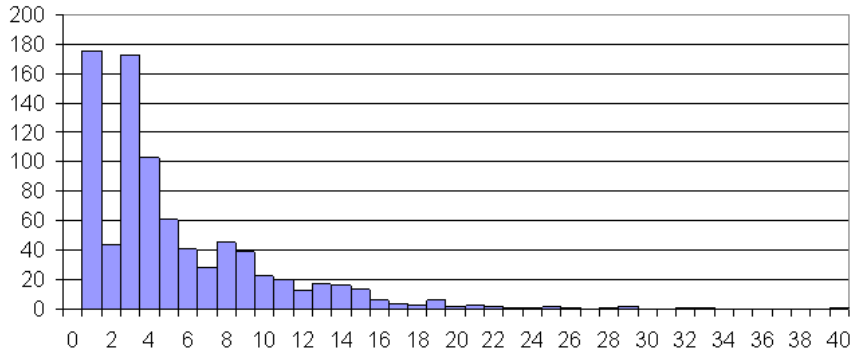

Figure 11 Distribution of length of 888 responses to the question shown in Figure 2 (without the wording shown in italics) in formative-only use.

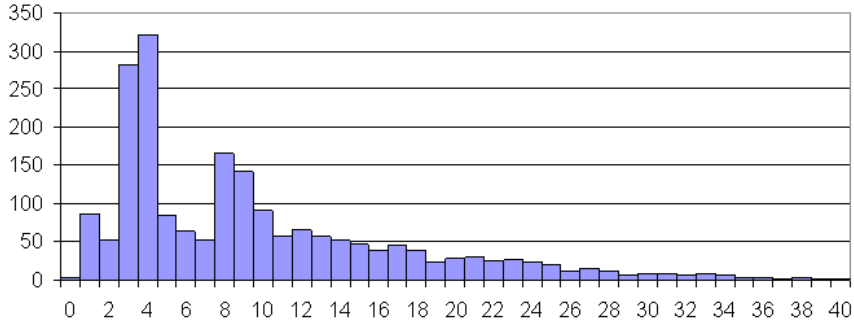

Figure 12 Distribution of length of 2057 responses to the question shown in Figure 2 (without the wording shown in italics) in summative use (Case 3).

Unfortunately some excessively long responses were received (up to several hundred words) to early summative versions of short-answer free text questions, and these frequently contained a correct answer within an incorrect one. Responses of this type are recognised as being the most difficult for computerised marking systems of any type to deal with and for this reason, from February 2009, a filter has been introduced to limit the length of responses to 20 words. Students who give a longer answer are told that their answer is too long and are given an extra attempt. The filter was initially accompanied by the text shown in Figure 2 (*You should give your answer as a short phrase or sentence. Answers of more than 20 words will not be accepted*).

The introduction of the filter and explanatory text reduced the number of students who added text to previous answers without thought to the sense of the response so produced.

It also dealt with the excessively long responses that were difficult to mark, and increased the number of students giving their responses as sentences. However, for all questions, the addition of the filter and explanatory text resulted in an overall *increase* in median response length (see the distribution shown in Figure 13).

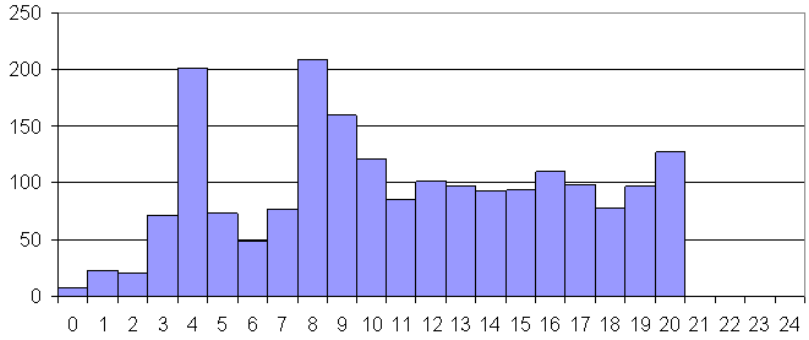

Figure 13 Distribution of length of 1991 responses to a question in summative use (Case 3) with filter and additional wording on the question.

A possible explanation of this effect is that more students were heeding the advice to give their answer as a sentence, now that this advice was given in the question. A less desirable explanation is that students were interpreting the advice to use no more than 20 words as indicating that they should be writing exactly or almost twenty words. From July 2009, the advice accompanying the filter has been changed to '*You should give your answer as a **short** phrase or sentence.*' The question shown in Figure 2 has not been used since then, but length distributions for other questions illustrate that the latest change of wording appears to have had the desired effect, reducing the median length and the number of responses that are exactly or just under 20 words in length.

Closer inspection of student responses to questions: (b) use of feedback

One of the anomalies of previous evaluations of e-assessment at the Open University and elsewhere is that the vast majority of students report finding the feedback provided on iCMA questions useful, but yet some are observed to make no use of it. In a first attempt to interpret evidence of actual behaviour, graphs have been plotted to show the number of incorrect responses that were unchanged at second and third attempt. Figure 14 illustrates the number of repeated responses *for the same question* in summative and formative-only use (for Case 2) and in diagnostic use (Case 5). In the figure, grey shading indicates blank responses, green shading indicates correct responses; red, orange or yellow shading indicates incorrect responses; an identical colour from first to second attempt or from second to third attempt indicates an unchanged response.

Thus it can be seen that a proportion of responses are left blank and in some circumstances almost half are left unchanged for a second and third attempt even when users have been told that their previous answer was incorrect. Not surprisingly, the proportion of blank and repeated responses varies markedly with mode of use, with very little of this behaviour in summative use, more in purely formative use and more still in diagnostic use. The proportion of responses that are blank and/or repeated also varies

considerably with question type; students are more likely to repeat responses when they cannot guess the answer (in a multiple choice or drag and drop question). It is possible that students are deliberately leaving responses blank or repeating them in order to receive the feedback provided at a later stage. Reasons for this behaviour will be explored further by interview.

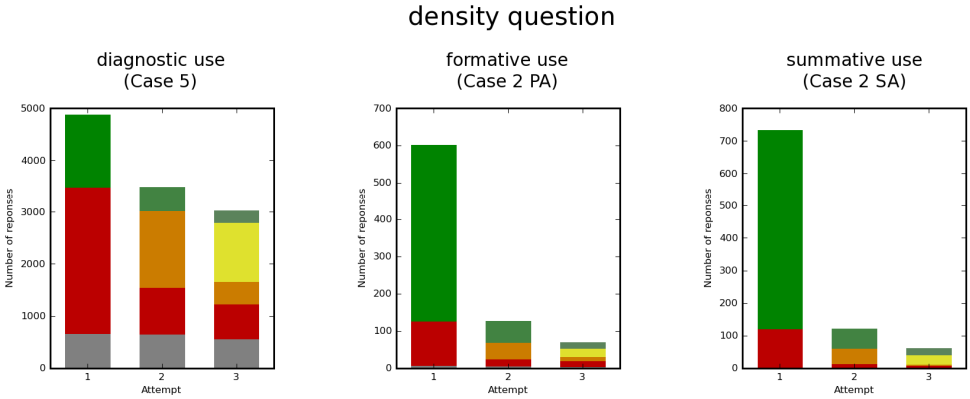

Figure 14 Blank and repeated responses for the same question in diagnostic, formative-only and summative use.

Conclusions

In general terms, students appear to engage with iCMAs in a deeper way when the questions carry some summative weight. However, in summative use, students become obsessed with the minutiae of the grading, as witnessed by many emails from students who believe – usually wrongly – that 'the computer has marked them incorrectly'. The use of thresholding or a 'carrot' (e.g. having similar questions in an unseen exam) may provide an alterative mechanism for encouraging students to engage with iCMA questions and so to learn from the feedback provided.

Things are not always as simple as you think they will be. At face value, use of the iCMAs in Case 3 and Case 2 SA should be similar, but this has been shown not to be the case. The differences appear to be entirely attributable to students' interpretation of what they have been told to do (in Case 2 SA they have been told to do the questions after studying the relevant part of the course; in Case 3 emphasis has been put on the importance of checking access to the iCMA in plenty of time, even if the student doesn't attempt any of the questions). Similarly, the increase in average length of responses to free text questions in response to the instruction that responses should be no more than 20 words in length, points towards a student interpretation that 'no more than 20 words' means 'nearly 20 words'. The fact that responses were more likely to be in complete sentences suggests that students may be more likely to read instructions when they are provided within the question, rather than hidden away in an introductory screen or in the course guide.

Ongoing work

The results reported here are just some of the outcomes from the quantitative analysis of the data captured when students attempt iCMAs. Other factors that are being investigated

include the length of time students spend in answering iCMA questions of various types and in different modes of use, and the order in which they answer the questions. The effect of variables such as start date, finish date, elapsed time and active time on performance will also be analysed, in order to investigate whether students who engage with iCMAs in different ways perform differently, for example, are early and late completers of iCMAs equally successful? Linked to this will be further investigation into whether students behave differently in different situations, for example, do students exhibit similar behaviour when they are revising to that when attempting an iCMA for the first time?

A comparison of iCMA scores, TMA scores and overall course performance will investigate whether iCMAs are a good predictor of success in other areas.

In autumn 2009, online questionnaires were sent to students on SM358 *The quantum world* (Case 4) and a range of level 1 courses, with the aim of discovering the reasons for some of the behaviour that has been observed, for example students leaving responses blank and repeating unchanged responses at second and third attempt. A subset of the students who completed the questionnaires will be interviewed.

From February 2010, four level 3 physics and astronomy courses (including SM358 *The quantum world*) will have a radically different assessment strategy, with all assignments being formative-only but thresholded i.e. students will have to complete a certain number of TMAs and iCMAs, where satisfactory completion of an iCMA will be deemed to be a score of 30% or more. A similar quantitative analysis of the responses to these courses' iCMA questions, and a comparison with findings from the current work (especially for SM358 in 2009 and for courses in which the iCMAs are in summative use) will be one of the mechanisms used to evaluate the impact of the change in assessment strategy.

For further information on this project, and updates, see
http://www.open.ac.uk/colmsct/projects/sejordan

Acknowledgements
The authors gratefully acknowledge the financial support of the UK Higher Education Funding Council via the Centre for Open Learning of Mathematics, Computing, Science and Technology (COLMSCT) and the Physics Innovations Centre for Excellence in Teaching and Learning (piCETL), the assistance of many people associated with COLMSCT and piCETL, especially Spencer Harben and Richard Jordan, and **t**he co-operation of the course teams involved in the investigation.

References
Boud, D. (2000) Sustainable assessment: rethinking assessment for the learning society, *Studies in Continuing Education*, 22(2), 151-167.

Butcher, P.G. (2006) OpenMark examples. Available online at http://www.open.ac.uk/openmarkexamples/index.shtml [accessed 14th Nov 2009].

Butcher, P.G. (2008) Online assessment at the Open University using open source software: Moodle, OpenMark and more. 12th International CAA Conference, Loughborough, UK, July 2008.

Buzzetto-More, N.A. and Alade, A. J. (2006) Best practices in e-assessment. *Journal of information technology education*, 5, 251-269.

Gibbs, G. (2006) Why assessment is changing, in C. Bryan and K.V.Clegg (eds) *Innovative Assessment in Higher Education*. London: Routledge, 11-22.

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

Hunter, A.G. (2008) Using an online formative assessment framework to enhance student engagement: a learning outcomes approach. Paper presented at the Improving Student Learning Symposium, Dublin, 3-5 Sept 2007.

JISC (2009) Accessed from http://www.jisc.ac.uk/whatwedo/programmes/elearning/assessment.aspx, 14th Nov 2009.

Jordan, S.E. and 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.

Jordan, S.E. (2007) The mathematical misconceptions of adult distance-learning science students. CETL-MSOR Conference, Loughborough, 11-12 September 2006. Maths, Stats and OR Network, 87-92.

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.

Rust, C. O'Donovan, B and Price, M (2005) A social constructivist assessment process model: how the research literature shows us this could be best practice. *Assessment & Evaluation in Higher Education*, 30(3), 231-240.

Sadler, D.R. (1989) Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119-144.

Scouller, K.M. and Prosser, M (1994) Students' experiences of studying for multiple choice question examinations. *Studies in Higher Education*, 19(3), 267-279.