INQUIRE: a case study in evaluating the potential of online MCQ tests in a discursive subject

Sophie Clarke\textsuperscript{a*}, Katharine Lindsay\textsuperscript{a}, Chris McKenna\textsuperscript{b} & Steve New\textsuperscript{b}

\textsuperscript{a}Academic Computing Development Team, University of Oxford, UK; \textsuperscript{b}Said Business School, University of Oxford, UK

There has been a wealth of investigation into the use of online multiple-choice questions as a means of summative assessment, however the research into the use of formative MCQs by the same mode of delivery still remains patchy. Similarly, research and implementation has been largely concentrated within the Sciences and Medicine rather than the more discursive subjects within the Humanities and Social Sciences. The INQUIRE (Interactive Questions Reinforcing Education) Evaluation Project was jointly conducted by two groups at the University of Oxford—the Saïd Business School and the Academic Computing Development Team to evaluate the use of online MCQs as a mechanism to reinforce and extend student learning. This initial study used a small set of highly focused MCQ tests that were designed to complement an introductory series of first-year undergraduate management lectures. MCQ is a simple and well-established technology, and hence the emphasis was very much on situating the tests within the student experience. The paper will cover how the online MCQs are intended to fit into the Oxford Undergraduate study agenda, and how a simple evaluation was executed and planned to investigate their usage and impact. The chosen method of evaluation was to combine focus groups with automated online methods of tracking, and the paper discusses the findings of both of these.

Introduction

Multiple Choice Question (MCQ) tests are one of the most widely used teaching tools, and have translated very successfully into the online environment; there are countless free and commercial products, and nearly every VLE has the built-in capacity to deliver online MCQ tests. Uptake has generally been dominated by the Sciences and Medicine, with less interest from the Humanities and Social Sciences.

*Corresponding author: Academic Computing Development Team, Oxford University Computing Services, 13 Banbury Rd, Oxford OX2 6NN, UK. Email: acdt@oucs.ox.ac.uk

ISSN 0968–7769 (print)/ISSN 1741–1629 (online)/04/030249–12
© 2004 Association for Learning Technology
DOI: 10.1080/0968776042000259564
This has been attributed to a belief that MCQ tests are not well suited to discursive subjects, and furthermore that they propagate in students the mindset that their education is about the retention and regurgitation of facts (McKenna, 2001b). The type of opinions expressed in McKenna (2001b) are typical of the reservations academics in these subjects have about MCQ tests. They are not convinced as to the suitability of MCQ tests for their subject area, and if MCQ tests can be of use then they are applicable only at a very unsophisticated level: ‘I do think that if you want to understand the basics of a course, it’s a good way of getting a basic answer’ (McKenna citing Weldon, 2001b).

This paper outlines a small-scale case study at the Saïd Business School, University of Oxford, which aimed to investigate further the potential for using online MCQ tests to support a discursive subject. The intention for these tests, or rather informal quizzes, was to look beyond the application of MCQ quizzes for testing the retention of numerous facts to see if they could be used to deepen student understanding, and to see how they might complement the traditional teaching methods used within the undergraduate course. The evaluation carried out concentrated on the impact of the formative assessment on student learning, as well as the student and staff preferences and attitudes where substantial work has been carried out in other studies (Charman, 1999).

**Background**

Undergraduate teaching at the Saïd Business School is very much in the model of the traditional Oxford tutorial system. Students have short periods of intense formal practitioner input in the form of tutorials, and these are complemented by lectures. Outside of this small amount of high-intensity formal input, students are expected to manage their own studies, including a large amount of directed and also self-discovered reading, writing essays for tutorials, and preparing for examinations. There are two aspects to this system that influenced the decision to investigate the potential for using MCQ tests further. Firstly, in common with most UK Higher Education institutions, there is an increasing pressure from students to provide as many learning resources as possible. In particular, students had requested additional ways in which to learn and judge their progress during periods of low contact time with their tutors—especially in the lead up to examinations. A second factor in the decision to investigate MCQ tests is the importance of the material covered in the tutorials and lectures. A great deal of emphasis is placed on the tutorials and lectures, and it is important that the ideas and concepts presented in them are grasped by the student. There is little or no opportunity for this to be done at a later date, as least not as part of the formal curriculum. Tutorials are viewed as high priority by both students and staff, whereas lectures are accorded lower priority by students when other commitments interfere, and so it was deemed especially important that the MCQ quizzes would support students who did not make the lecture.

The reasons above highlight why the Business School decided to trial online formative assessment as a means of providing students with an additional resource
to support the cycle of lectures, essay-writing and tutorials. The pilot implementation of MCQ quizzes to support the lectures is discussed here, and in particular the results of the evaluation carried out to investigate their first implementation in the undergraduate curriculum. The scope of the pilot involved the use of online MCQ quizzes to support the ‘Introduction to Management’ course, taken by all undergraduate students. The study evaluated the use of an online MCQ test alongside the lectures to investigate students’ use of the resources. The aim was to develop further quizzes to support more of the undergraduate curriculum if they proved successful.

The MCQ quizzes had several key characteristics. The questions themselves were carefully crafted to address the educational objectives outlined below. Each individual potential answer for each MCQ question needed to be able to have its own feedback, and in both the questions and answers there was a requirement for the ability to include hyperlinks and images. A variety of MCQ software was available at the University, and software choice was regarded as relatively insignificant at the start of the project. However, the feedback and HTML inclusion requirements dramatically limited the choice of suitable software. The solution finally chosen was Quia, an online activity creation tool hosted on the service providers web site. Quia fulfilled the aforementioned requirements, and was also suitably easy to use for non-technical question authors.

Aims and objectives

There were four areas in which it was thought that the MCQ quizzes could help students, and the pilot tests were carefully written with these aims in mind. Investigating the success in fulfilling these aims was used as the basis for choosing an evaluation methodology. The four aims it was hoped that the MCQ quizzes could address are outlined below.

The reinforcement of learning from other pedagogic elements

As previously discussed, the pressure for lectures and tutorials to deliver effectively is high, but there are many factors that can contribute to whether or not a student succeeds with a particular subject. Lectures can vary in quality across the course, and are more likely to fall by the wayside completely in terms of student’s priorities when other commitments press, such as assignment deadline and extra-curricular activities. Logistically the structure of the course means that there can be many months between the last formal teaching input on a particular topic and the exam.

It was envisioned that the MCQ quizzes would provide students with a further opportunity to cover the lecture material when and where it suited them. For students who did not make the original lecture, they would have extra help in comprehending the lecture handout, and a way of measuring their own level of understanding.
Cementing students’ understanding

MCQ quizzes have a role in terms of testing that students have grasped the key facts and ideas. Additionally, this is a practical safeguard as the lectures are designed to challenge the students rather than merely transmit information. In addition to covering the core concepts, MCQ questions can be used to cement students’ understanding of the more subtle points in the lectures, particularly through the use of the carefully constructed feedback.

Deepening students’ knowledge

The MCQ quizzes provide an excellent opportunity to present entirely new material to students, leading them beyond their reading lists. For the more curious students, they can provide some framing to new areas that there might not be time to cover in the tutorials, and can guide the student towards credible resources that they may not otherwise have found. Necessary to this end is the ability to link out to online resources (the links made ranged from company web sites to journal articles in the department electronic library provision).

Framing educational expectations

MCQ quizzes provide another way for course leaders to convey to students (explicitly and implicitly) what is expected of them. Many first-year undergraduate students are surprised and daunted to find that they are expected to read complex journal articles and whole books. The MCQ quizzes can reinforce the message of what they are expected to read by basing questions on those materials and by linking out to a wider variety of literature than is present on the reading lists. In general, students can be pointed in particular directions without being obviously spoon-fed.

Example question and answer with feedback

In the lecture, you learned that even in recent years the shift to working in large organizations continues. What do you think that this has meant for the number of hours that people work?

In other words, do you believe that it is true or false that working for larger organisations means that people work longer hours than they would have if they were working for smaller companies or for themselves?

(Hint: You might want to consult this study on long term employment trends: http://phe.rockefeller.edu/work_less/)

- True
- False

Answer: False

Feedback: For a critical view of the conclusion that people are working fewer total hours, see Juliet Schor’s book The Overworked American.

There were several potential pitfalls identified with respect to the MCQ quizzes, and attempts were made to avoid these, particularly in the writing of the questions. There
is a danger that by picking out particular areas (either deliberately or inadvertently), the quizzes could send misleading clues to students about what is and isn’t important. This is exacerbated by the students’ tendency to be very strategic and exam-focused when considering how best to spend their study time. A second potential problem to avoid was writing overly simple questions with black or white answers, or indeed the opposite issue of inadvertently adding subtle nuances that some students may wrongly pick-up. To avoid these pitfalls questions were authored collaboratively and tested before they were made available to the students.

**Methodology**

This section outlines the evaluation methodology used to consider the usage of the six pilot MCQ quizzes over the course of a term. The aim of using evaluation techniques was to assess if the quizzes delivered any or all of the four perceived aims outlined above. This type of subjective evaluation is notoriously difficult to design and carry out. The expectation for the evaluation was to get an initial indication of how well the pilot tests had worked. From there any particularly promising, or interesting, aspects could be identified to be investigated in greater depth and with more resources (possibly with a larger group of students over more than one course).

A range of evaluation methods were considered, and each assessed against its potential to illuminate how well the MCQ quizzes had addressed the four primary aims. A focus on methods with a proven track record in educational research and the evaluation of technology-based resources led to a shortlist of six possibilities: system logs, automated tracking of link following, confidence logs, 1-minute papers, video observation, focus groups (Angelo & Cross, 1993; Harvey, 1998). The shortlist was then considered again to judge which methods could address multiple aims at once. Between the chosen methods they would be required to address all of the four aims, and to provide a combination of qualitative and quantitative data.

Access logs, the tracking of link following and focus groups were chosen as the most appropriate. Techniques that attempted to quantify student knowledge in some way (confidence logs and 1-minute papers) were discarded on the grounds of difficulty of application within the resources available, and also because of the problems inherent in using these methods to judge improvement against the stated aims. Video observation of students was also decided against as the focus was more towards the cumulative result of accessing all six quizzes rather than in analysing the specific answers chosen in any one sitting. Similarly, there was no particular concern with software usability at this stage, an area in which video footage would be particularly applicable.

Access logs provide an unobtrusive method of mapping the use patterns of individuals. They give a useful picture of how the quizzes were used in relation to the timing of other aspects of the course (lectures and tutorials), and also provide quantitative data to back up the focus groups. There are data protection issues surrounding this,
and students were asked to give their consent to tracking using anonymous but distinct usernames, i.e. student1, student2. The specifics logged by this method were the details as to the date, time and frequency with which each username accessed each test.

Tracking of link following was used to measure whether or not the students moved beyond the boundaries of their recommended reading and used the hyperlinks embedded in the questions, feedback and answers. Code was embedded directly within each hyperlink, this took the student to the page advertised but via a script that recorded the following of that link in a database. Tracking done in this way is extremely quick and is not visually obvious or intrusive to the student. However, ethics again came into play and student consent was necessary.

Finally, focus groups were chosen as the best method of getting the students opinions on the quizzes and judging the effect on their learning and understanding (Morgan, 1988). The focus groups were chaired by a course leader familiar with the curriculum with some lines of investigation intended to assess student understanding of the lecture topics. To help ascertain to some degree the effects of the MCQ quizzes as opposed to the learning bestowed by the course as a whole, it was decided to use three sample groups. Because the quiz aims were as yet unproven, and because they would be made available to all students at a later date if beneficial, it was considered to be an ethically acceptable strategy. A self-selecting group of thirty student volunteers were involved in the evaluation. The groups are described below:

The control group

The control group students went through course in the same way as those not involved in the trial. They attended the lectures and tutorials in the usual way, but had no access to the MCQ quizzes.

The paper group

This group of students completed the lectures and tutorials. In addition, after each lecture they were given a printed version of the MCQ test, with answers provided on a separate sheet. Online material referenced in the questions was given as printed URLs, and references to journal articles and books were given in the form of traditional written citations.

The online group

This group of students completed the lectures and tutorials. After each lecture, an online MCQ test was released. Students were given immediate feedback after answering each question, URLs were provided as hyperlinks and journal articles and books were also provided as hyperlinks wherever possible. Each student had an individual but anonymous account as described earlier.
Results analysis

The investigation took place over six weeks, with the staggered release of six sets of quizzes coinciding with the six introductory lectures to the course. Throughout this time the logs and tracking were analysed and at the end of the six-week period the students were brought together in their individual focus groups, not only to discuss the impact of the quizzes, but also for analysis of their level of knowledge and understanding concerning the concepts and ideas of the lectures that they had been to. The results that emerged gave a valuable insight into how a resource such as formative online multiple-choice questions can be used to reinforce lectures and enhance student learning. What follows is a results analysis of the focus groups, tracking and logs in relation to the four aims posed, and a discussion of other relevant findings.

The reinforcement of learning from other pedagogic elements

To monitor whether the quizzes were successful as a pedagogy to reinforce learning, the usage of the quizzes was analysed as well as students’ attitudes towards using them as an additional study resource. The paper group voiced their reluctance to use the quizzes. It was inconvenient to pick up the quiz sheets after their lectures and they were generally put to the bottom of the pile of paper accumulating on their desks. Being delivered on paper was not appealing as the quizzes lacked a high level of interaction, students found them bland and boring and none admitted to going back to them after completing them the once, although some filed them away for ‘revision purposes’.

For the online quizzes the collected logs were examined to show usage. Whilst such an analytical method can only provide a broad overview of the students’ use patterns, the statistics collected provided some valuable information. In some individual cases students accessed all the quizzes for the first time a couple of days before the focus group session. Here it could be argued that their behaviour was affected by the knowledge that they were being tracked and of the impending focus group, their attention drawn more to the act of evaluation rather than the act of learning (Faulkner, 2000, p. 171). For the remainder of the students the statistics showed that the quizzes were subjected to light usage; 100% of these students accessed all of the quizzes, but only 40% accessed them more than once. During the focus group session the general consensus was that the quizzes were a valuable resource for many students:

Student C: … some people use different libraries, different readings and different newspapers, I think all of us—we all like choose how we study and stuff so it’s, I think, it’s valuable to have this as an option as well. Cos as I said before, it’s not for everyone, but for sure it’s valuable for a lot [of students].

The lack of repeated use of the quizzes was largely due to the difficulty students had in perceiving how they would form a cemented part of their studies, as it did not fit into the rigorous essay writing and examination process. Some students did go back
and redo certain quizzes where they felt they needed to improve. However, like the
paper group, the majority left it at one attempt with the forethought that such a
resource would be more useful in the period leading up to examinations. With this in
mind students commented that they would like the quizzes to be ‘a little more based
on the exam’ rather than being based on the lectures.

The times at which the students accessed the tests showed a strong pattern of late-
night quizzing. The majority of log-in times occurred between 8pm and 12am in the
evening. On questioning the students in the focus group it became clear that the
process of taking a short quiz provided a work-related distraction for them whilst they
were engaged in another activity, such as preparing for a tutorial or waiting to meet
friends:

Student A: It’s like something you can do [so] that you feel like you’re doing something
useful …
Student B: It’s kind of like in-between time … after dinner doing work and maybe you go
out later to meet your friends or something.

Cementing students’ understanding

The understanding of the key concepts covered during the lectures differed between
the three groups. The control group and the paper group had extremely patchy
retention of the details of the lectures. They had yet to grasp the key concepts and
facts that had been delivered and generally had to bounce off each other when
recalling lecture topics: ‘I don’t really remember much of it but when people start
saying stuff it starts coming back’. The paper group thought that the quizzes were
‘good as a memory trigger’, but there was little evidence within the focus group
sessions that they had remembered significantly more than their peers. It was clear
that the students’ knowledge was built primarily through essay writing and tutorials
rather than analysing and going back to the lecture content. In the lectures they
sought specific facts and ideas which would help them with that week’s essay: ‘he’s
[the lecturer] probably going to be mentioning something that is very relevant to your
essay … and then you want to capture certain points’.

The online group were noticeably more forward in their focus group when
discussing topics that specifically related to the lecture content and more readily
recalled subtle points that had been addressed in the lectures, as well as key examples
and facts. The group all agreed that the online quizzes had provided a way of helping
them remember the lectures:

Student A: I think it’s a good thing to go back, sit down and answer questions, because if
you answer questions you think about it more than when you just sit down and read the
notes again.
Student B: And then sometimes you miss things, at lectures, you know you’re not always
concentrating. So, you know, when you read the questions … they help you think about it
more and, more in-depth.
Student C: Yeah, I think it’s good because, it’s good to think about it more, and it’s the
right kind of style, it is for learning like. I think he raises key points that he wants you to
get out of the lecture, impressions.
Student C: Sometimes I felt that.
Student D: Yeah, it kind of triggered your memory about, I waited about a week, two until after the lectures to do it ... I remembered it better in my mind, doing that. Cos, I think if I'd done it straight away, it would have kind of all been one thing and I think I would have just forgotten it all after a while.

Interestingly, the images used in the quizzes provided a particularly useful hook with the lecture material, jogging their memories and reinforcing ideas and facts that had been discussed.

Student A: I would use as many pictures, illustrations as you did, because they can sometimes just give you that little bit that makes you remember them. Cos when you see a picture, a photograph of Thomas Edison, you think, oh hang on, yeah, General Electrics, because it was on the slide. But when they weren't on the slides, they sort of confuse you a bit, because you think I must have seen that before, but you haven’t.

**Deepening students’ knowledge**

To achieve a deeper understanding of the course students were given links to journal articles and company websites. In the paper-based quizzes these links were written down as URLs or references to hard copy books and journals. The students found typing out the URLs ‘annoying’ and if they were not actually sat at a computer at the time of taking the quiz they were unlikely to follow them up. Likewise, the references to hard copy books and journals were not pursued as they felt they were reading enough with the set tutorial texts they were given each week: ‘I don’t do any more reading than I have to’. As a result the students found it difficult linking the lecture topics that they could remember to relevant wider material outside of their reading lists.

The link tracking showed that the students taking the online quizzes did follow the links to the outside resources. Interestingly the students refuted this—denying that they had read the articles and company profiles:

Interviewer: ... for the ones [questions] which had the links, which could take you off and give you something more, long to read, did anyone look at those?
Student E: No, I didn’t.
Student C: I read some, I just didn’t include them.
Student A: Depends how long they were. Sometimes I read through them but sometimes I thought, ooh that’s too long, for the moment.

It is clear that many students just clicked-and-glanced at the wider reading materials. This, they considered, did not constitute learning as there was no depth of reading or interpretation. However, it was evident that this very mundane practice made the students acknowledge that there was a wider pool of resources that they could, and were expected to, consult within their study practices. Links to scholarly articles that the students felt relevant to their essays received a greater number of hits than others and the company websites. Many students commented that they had added these to their browser ‘favourites’ folder to read at a ‘later date’.
Framing educational expectations

Through the focus group sessions it was clear that students were starting to take on the expectations of the department. There was little difference between the three groups indicating that the quizzes were a precarious factor in framing the students’ educational expectations. There was a slight difference in the online focus group session where there was a strong general consensus that they were expected to read much more than they were already doing, perhaps as a result of following the links placed to deepen their knowledge.

Further findings of the evaluation

All three focus group sessions brought to light some interesting points concerning the nature of the students’ study patterns at the university. Although not directly related to the research questions of this study, these highlighted some pedagogic concerns which are relevant to the development of online learning tools. Firstly, it was clear that undergraduate students study habits were entirely governed by the tutorial and essay writing timetable—each academic week revolved around the process of reading and writing for the next tutorial. This rigid study routine gives students little incentive to read any wider than the set papers for their forthcoming tutorials, rarely consulting further reading lists or searching for wider materials themselves.

Secondly, it became evident that the lecture handout played a talismanic role in the students study process. The students were immensely grateful for the lecture handout. For them it holds a vitally important place in their resources for study, representing a walkthrough of the lecture. However, to the lecturer it only really offers an overview from which they expect their students to expand their knowledge through further reading and discussion. Because the lectures were engaging and entertaining, students took few, if any, notes and tended to sit back and enjoy the class; they felt they could use the handout later as a reference point. Interestingly, although the handout appears to be an important resource, students can remember very little of the content of their lectures when quizzed. Filed away and retrieved nearer the examination period it provides an outline of what needs to be regurgitated for a particular topic.

The rigorous structure of lecture-tutorial-examination presents to the advocator of e-learning a problem of materiality. When given a set reading list and lecture handout students are inclined to do little more than follow the lines of investigation that these both represent. This structure is not conducive to wider thinking and obtaining a deeper knowledge. Additionally, learners want to make sure that the skills and capabilities they acquire through using study resources are relevant to their summative course marks and it is not clear to them how optional e-learning such as formative quizzes can contribute to their degree.

Conclusions

Lectures are a fundamental part of the university teaching system – they provide an overview of a course’s topics and key ideas, and guide students towards areas of
further research and investigation. However, in a university system that also includes high quality tutorials it is often the case that lectures are put aside when there are more pressing academic activities and events in a student’s life. From the INQUIRE evaluation it is strongly indicated that reinforcing the content of the lectures through formative assessment can act to cement students’ understanding of key concepts and ideas. The value clearly lies in the ‘jogging’ effect that the quizzes had on the students’ memories, and fundamentally the mode of delivery influenced how effective this was. Good practice in university teaching places a high emphasis on the entertainment value of a lecture with its quality of being engaging and thus memorable, the implication being that memorable lectures are more effective in terms of student learning. We need to realize that the same is true for learning resources.

This evaluation project, like others before it (e.g. Mansen & Haak, 1996; Martin et al., 1996; Charman & Elsmes, 1998a) concerned chiefly the attitudes and preferences of the students towards using formative online assessment. It also attempted to address the impact of this type of assessment on student learning through comparing the knowledge and understanding of the three student groups. The evaluation of the effectiveness of learning technology is particularly complex and is fraught with difficulties: the success will vary, depending on the scope of previous knowledge, attitudes, and conceptions which particular learners bring to the learning situation, and the larger context in which the learning situation is embedded (Jackson, 1998). The qualitative results of this evaluation suggest that online MCQs can reinforce learning from other pedagogic elements and cement student understanding in subjects that were not traditionally thought to lend themselves to such a model (McKenna, 2001b; Leon, 2002). However, to clearly establish the positive impact of this method of formative assessment there must be further investigation of a quantifiable nature. The most valid option in the current educational system is to chart examination performance, as Charman states ‘the only real way to assess this quantitatively is to compare performances in summative assessments’ (Charman, 1999). It will be highly interesting to observe over the next few years if there is a marked difference in the module examination results attributable to the online quizzes.¹

The use of formative online MCQs could potentially offer measurable benefits in supporting humanities and social science students in their study. However, there is still much to do if this is to become fully integrated within the study experience. It is still widely perceived both by education establishments and students that all forms of assessment are a means to measure knowledge and are inextricable from certification. The aims of these quizzes, however was to compliment traditional teaching and deepen understanding rather than to assess competence in a subject. There is still much work to do if there is to be a shift in our assessment paradigms (Birenbaum 1996; Sambell et al., 1999) and the benefits of formative assessment can be realized and integrated into the rigid study structures posed by many higher education systems. Although this was a small-scale evaluation it has laid some useful foundations for further research in this area.
Notes

1. For similar studies that have compared exam performance, see Boyle et al., 1997; Paul & Boyle, 1998; and Charman & Elmes, 1998c,d.

References


