

AUTOMATED MULTIPLE-CHOICE TESTING FOR SUMMATIVE ASSESSMENT: WHAT DO STUDENTS THINK?

Pete Bridge, Rob Appleyard, Rob Wilson

Sheffield Hallam University, Sheffield, UK

p.bridge@shu.ac.uk, r.m.appleyard@shu.ac.uk, r.wilson@shu.ac.uk

ABSTRACT

This paper reports undergraduate student feedback contrasting conventional “Long-answer” examinations with automated multiple-choice question (MCQ) assessment. Feedback was gathered after students had undertaken formative MCQ assessments as a revision aid. Feedback was generally supportive of MCQ summative tests, with 74% expressing a preference for the new format. The examination conditions were preferred by 69% of students. Results indicate that students are in favour of the use of automated MCQ assessment. All topics can be reliably and validly assessed with an associated time saving of over 16 hours. The need for rigorous question and answer construction has been highlighted, but so long as sufficient care is taken at that preliminary stage, the overall benefits of the format outweigh the problems.

Keywords: Summative assessment, E-learning, Multiple Choice

INTRODUCTION

Increasing student numbers are presenting a range of challenges to Higher Education institutions and while delivery can generally accommodate this increase, the same cannot be said for assessment. The marking alone represents around 150 hours per person in the Sheffield Hallam University Radiotherapy and Oncology team. A potentially useful way of reducing this workload on staff is by using automated marking systems. Although there is considerable use of automated tests across the Radiotherapy course, they are currently performed as formative assessments. A previous study (Bridge and Appleyard, In Press) has demonstrated the usefulness of electronic submission of summative assessment via a virtual learning environment (VLE) and the next logical step in ensuring that its full capabilities are utilised is the development of summative assessment within the system. MCQ testing has been used successfully for 20 years (Caruano, 1999). Many studies have validated the use of multiple choice questions (MCQ) as part-assessment of undergraduate health students (Fullerton et al, 1997; Hammond et al, 1998; Brady, 2005). MCQ testing is efficient for large numbers of students and can be both a reliable and valid means of covering a broad range of content, as detailed in MCoubrie’s (2004) literature review. It is this efficiency and ability to assess a large number of topics reliably that makes automated MCQ testing an attractive option for content-rich modules. A pilot study was devised to compare student experiences of automated MQC and conventional unseen, written examinations in order to gauge the benefits and challenges associated with the change. This paper presents the student feedback from the study.

METHOD

A pilot study was developed with the aim of determining the feasibility of using automated multiple choice question (MCQ) tests for summative assessment for the Level 4 (Year 1 Undergraduate) “Radiotherapy Physics and Equipment” module. Students undertook 2, 25-question MCQ tests (1 per semester) via the VLE as a formative revision aid. Although formative, the assessment was performed under examination conditions, including the use of invigilation, and was strictly “closed-book”.

MCQ Construction

Each question comprised a question “stem”, a correct answer and 4 incorrect “distracters”. Questions were constructed in accordance with Holsgrove’s (1992) five rules for writing good questions. These can be followed by ensuring questions are simple, unambiguous and grammatically correct. Trick questions and negative stems are also to be avoided (Holsgrove, 1992). Current question construction follows the comprehensive guidelines suggested in Tarrant et al’s (2006) Appendix A. Questions took up to 40 minutes to construct, although Farley (1989) suggested that 1 hour is needed per question. Questions were subjected to review by a team of lecturers as suggested by Race and Brown (2001). A mixture of factual recall, interpretation and calculation questions was used to test those levels of Bloom’s taxonomy that applied to the learning outcomes for the module (Bloom, 1971). Most of the learning outcomes at this level are related to knowledge, understanding and application (the first 3 levels of Bloom’s taxonomy). More able students are capable of demonstrating analysis and a few questions tested this ability to aid with stratification. All students were asked the same questions, although these were presented in a random order to minimise the chances of copying from adjacent monitors. The answer stems were also presented in a random order to prevent the uneven distribution of correct answers alluded to by Tarrant et al (2006). Other

areas of the VLE were disabled to prevent students browsing the module content.

Procedure

Forty-two students completed the Semester 1 assessment and survey but only twenty-seven attended the Semester 2 event (See Table 1). The conventional examination comprised a 2-hour unseen “long-answer” paper in Semester 2. Candidates were presented with 8 questions, out of which they were required to answer 4. A simple online questionnaire was used immediately after the tests to collect student feedback via a mixture of 5-point Likert responses and open questions. Student feedback was obtained regarding their perceptions of MCQ examinations as compared with conventional “long-answer” examinations.

RESULTS

Time taken to complete the tests varied from 15 minutes to 50 minutes (with a mean time of 27). Feedback was generally supportive of the introduction of MCQ summative tests, with 76% of students deeming it to be a fair method of examination and 74% expressing a preference for the new question format. See Table 1 for a summary of the full results. The examination conditions were preferred to conventional conditions by 69% of students. Further study is ongoing to determine a measure of correlation between VLE-based MCQ and paper-based long-answer performances.

Student Feedback – Examination Conditions

Table 1 :- Student Questionnaire Feedback

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
MCQs are a fair method of examination of this topic	31%	45%	12%	10%	2%
I prefer answering MCQs to Long-answer questions	43%	31%	14%	10%	2%
I preferred the exam conditions for the MCQ test	33%	36%	17%	12%	2%
This module should be assessed with a VLE MCQ	31%	36%	12%	19%	2%
This assessment was too easy	0%	5%	31%	40%	24%

There were many

comments relating to the less stressful nature of the examination:

- “*Make me panic less about it!*”
- “*Feel that the environment is less stressful than sitting in an examination room.*”

These could be slightly misleading because the tests were not part of the students’ summative assessment and so lacked the importance of the conventional examination. Although examination conditions, including invigilation, were maintained, these comments suggested that the environment was less intimidating. Three students indicated a dislike of reading from computer screens:

- “*Don't like staring for long periods at computer screens, it gives me headaches.*”

Random question and answer orders were used to reduce the possibility of glimpsing the correct answer and students were made aware of this. It was felt that the students were not unfairly disadvantaged by the use of computers for assessment. This is supported by Lee and Weerakoon’s (2001) study comparing paper and pen tests with MCQs that found a low level of computer anxiety among biomedical science students despite little previous computer experience.

Student Feedback – Writing skills

Another common theme highlighted student concerns about their writing skills. Comments suggested that many students liked not having to worry about spelling or compiling long answers:

- “*You get credit for your understanding and knowledge and it stops you from phrasing questions incorrectly and thus losing marks.*”
- “*I worry less about this type of examination, since my spelling is no [sic] the greatest and therefore does not affect me in this*

This must be contrasted, however, with the later comments regarding the constrictive nature of the MCQ by students who prefer to explain their answers:

- *“There is either a right or wrong answer and no where to explain the knowledge of the various processes”*

Student Feedback – Question Compilation

Some students reported that MCQs made it easier to understand the question and thus ensured that they answered the question correctly.

- *“The answer is there in front of you and as long as you have the basic knowledge you should be able to work out which answer is correct.”*

Although students reported that the questions were less open to misinterpretation than long-answer style questions, there were some comments relating to similar answers causing confusion:

- *“Too many similar choices. You think you know the answer and then the multiple options can sometimes confuse or put you off what you originally thought”*

What this feedback does demonstrate is the importance of phrasing examination questions in a clear and unambiguous manner, irrespective of the format. With the MCQ format, attention must be paid to the different options presented, particularly with regard to avoiding ambiguity or similar answers (Holsgrove, 1992). Students had clearly struggled in some cases to distinguish between different options and felt that this had disadvantaged them. However, the ability to make this distinction will highlight the stronger students. So long as only 1 answer can be correct, or the question is phrased so as to allow the student to choose the most appropriate answer, this type of question is essential.

Additionally some students clearly missed having the opportunity to explain their answers and felt that they could have gained more marks by expanding their answers. This comment does indicate rather that students are used to having their writing skills assessed. Since it is mainly knowledge and understanding that is assessed in this module, the format remains appropriate. Although this could be interpreted as more able students feeling restricted by this type of examination, the higher level questions would have allowed them to demonstrate their evaluative skills, essential for discrimination between high and low ability students.

Care taken during compilation is clearly essential for the success of MCQ examinations. The ease with which prompts or ambiguity can arise was evidenced by Tarrant et al’s (2006) paper that discovered item writing flaws in almost half the MCQs tested.

Student Feedback – Speed of Feedback

Some students commented that they had enjoyed receiving the immediate feedback of their score rather than waiting for weeks to discover if they had passed. If negative marking is to be applied, however, this may not be as useful an indicator for the students and they need to be made aware that their score could drop, depending on how many questions they got wrong or omitted.

Student Feedback – Guessing

Some students felt that the MCQs were easier than conventional examination questions, although neither the scores nor the answers to the specific questionnaire section (Table 1) reflect this. The fact that the answer is presented to the student and can act as a trigger was a common perceived benefit, although this does reveal that the knowledge and understanding that was being tested was actually there. Some students evidently perceived the possibility of guessing as a potential problem.

- *“I suppose you could guess at some of the questions, but that would not be a true reflection of your knowledge.”*

Negative marking is controversial, but if used correctly can negate the effect of a wild guess. The effect is to cancel out the marks gained by guessing by weighting the marks in favour of the correct answers (Hammond et al 1998). “Intelligent guesses” that use existing knowledge to eliminate several answers have an improved chance of gaining marks, but this allows marks to be credited for partial knowledge. Other options that could counteract the guessing include normalisation or introducing elements of confidence assessment into the questions as described by Gardner-Medwin.(1995).

DISCUSSION

Question Construction

The student feedback highlighted several benefits and potential problems associated with the use of MCQ tests delivered via a VLE for summative assessment. Many of the problems related to construction and phrasing of questions and can be resolved by maintaining a rigorous and evidence-based approach to compiling MCQ questions. Tarrant et al (2006) highlighted the difficulties associated with question compilation, discovering that 46.2% of baccalaureate nursing MCQ assessments were poorly written, with more than 90% failing to demonstrate the higher cognitive skills proposed by Bloom in 1971. This is a common criticism of MCQ use for undergraduate modules. Higher order skills can be tested with MCQs (McCoubrie, 2004) but for this knowledge and comprehension-based module, these were used sparingly in order to assist with stratification rather than to provide a purely evaluation-based assessment. Clearly the care taken with writing conventional examination questions also needs to be applied to the MCQ format, but attention needs paying to distracter answer construction, question phrasing and cognitive level. There were further issues raised by this study that relate to the University perspective, which will be discussed now.

Impact on Workload

As previously alluded to, the main advantage of automated marking is the reduced marking burden of academic staff. In addition to the obvious reduction in actual marking, there are associated time-saving benefits. There is no requirement to construct an "answer key" (Schuwirth & Van der Vleuten, 2004) as there would be for essay-style assessments. Marks from the assessment can easily be uploaded into spreadsheet software at speed with no transcription errors. The burden on external moderators will also be reduced if they are able to approve the questions and answers prior to the exam. Subsequent moderation should not be needed due to the inherent accuracy and objectivity of the system. There is a clear financial and environmental benefit afforded by reducing the printing burden. Although invigilation will be required, there is no marking or administrative burden. The only increase in workload relates to the significant time needed to design reliable and valid MCQs with plausible distracters (Brady, 2005). For this study, the compilation took up to 40 minutes per question (up to 1000 minutes). Long-answer exam compilation (including answer-key) can take a similar length of time per question, leading to a total time commitment of 320 minutes. Previous experience of marking 4 long-answer scripts suggests that marking takes about 10 minutes per answer (40 minutes per student). For a cohort of 42 students the marking commitment equals 1680 minutes. This suggests a time-saving of over 1000 minutes for the module. Further time-saving can be achieved if questions are reused in future question banks.

Impact on Quality

Moving to automated marking of multiple choice questions has the potential to ensure strict marking accuracy and parity across the student cohort. The reliability and validity of the exam, however, is strongly dependent on devout attention to the question construction. The change in format offers an additional benefit to the assessment. With the current examination system offering a choice of questions, students can 'opt out' of unpopular topics and possibly miss out on essential learning. By asking many small questions on a range of topics, students are encouraged to engage with all of the module content. The large number of questions that can be asked can increase the reliability per hour when compared to conventional long-answer formats (Schuwirth & Van der Vleuten, 2004). The reduced time taken to answer can reduce the candidate fatigue that is often associated with long-answer examinations. MCQs have the potential to ensure that knowledge rather than writing skills or stamina is being tested. For this underpinning first-year module, multiple choice questions offer reassurance that students must engage with all topics, thus matching the assessment format to the purpose as recommended by Crossley et al (2002) in a paper defining principles of good assessment design. Despite the perceived usefulness of MCQs for this module, the value of a range of different assessment strategies across the course must be emphasised (Fischer et al, 2005). Different modules and levels will contain a diverse mix of content and skills and a range of assessments will not only reflect this diversity but also appeal to different personality traits across the student cohort (Chamarro-Premuzic et al, 2005).

One possible reason for student satisfaction with the new format is a recent finding (Chamarro-Premuzic et al, 2005) that MCQs are disliked by individuals who express the personality trait "Open" (those most likely to study humanities and arts). Since radiotherapy is primarily a science-based discipline, the student profile is not likely to include many "Open" individuals. Further study into the personality traits of the student cohort is needed to confirm this supposition.

Student Support Issues

A potential problem for students relates to their unfamiliarity with the MCQ format. Technique can play an important part in success at MCQ examinations with 74% of "educated guesses" being correct (Hammond et al, 1998). Students clearly need some degree of coaching on MCQ technique in order to ensure they do not penalise themselves either by risking a truly "wild guess" or by not taking an educated guess. Another aspect of exam technique was highlighted by Fischer et al, (2005) who demonstrated that students should not automatically go with their "gut instinct" but should be encouraged to change answers

they previously had doubts over. The authors found that 55% of changes were from an incorrect answer to a correct one and only 25% were from a correct to an incorrect. Student technique can be enhanced throughout the module by providing plenty of practice as formative topic tests. All students are expected to integrate with the VLE on the course and a previous study (Bridge and Appleyard, In Press) found that 94% of students expressed their skills with the VLE to be "OK" or better.

Potential for Cheating

The real possibility of using the Internet or even the VLE course materials during the examination must be acknowledged, although the time taken to browse for the answer would undoubtedly disadvantage the student. Course materials can be made unavailable for the duration of the examination, but measures may need to be taken to prevent students from accessing the internet. Delivering VLE assessments as an "offline" application may be one appropriate solution. Copying from adjacent computer screens can easily be avoided by ensuring that not only questions but also potential answers are presented in a random order. Strict invigilation was used successfully in this pilot project and this will obviously need to be applied in future examinations.

CONCLUSION

The results of this study indicate that a VLE-supported MCQ assessment tool is a well-accepted assessment format for a content-rich first year undergraduate module. The whole range of topics can be assessed with a range of questions testing knowledge and understanding as well as application and evaluation. Students are generally supportive of the format and it is estimated that the introduction of this format to a 42-strong student cohort would save over 1000 minutes of assessment-related tasks. There is a need for student coaching in examination technique to ensure they are not disadvantaged. The need for rigorous question and answer construction has been highlighted, but so long as sufficient care is taken at that preliminary stage, the overall benefits of the format outweigh the problems.

REFERENCES

- Bloom, B. (1971) *Taxonomy of Educational Objectives: The Classification of Educational Goals: Handbook 1: Cognitive Domain* New York: David Mackay
- Brady, A. (2005) Assessment of Learning with Multiple-Choice Questions, *Nurse Education in Practice*, 5, 238-242
- Bridge, P. & Appleyard, R.M. (In Press) A comparison of electronic and paper-based assignment submission and feedback, *British Journal of Educational Technology*
- Caruano, RM. (1999) An Historical Overview of Standardised Educational Testing. Available online at: <http://www.gwu.edu/~gjackson/caruano.PDF> (Accessed 6th March 2007)
- Chamarro-Premuzic, T., Furnham, A., Dissou, G. & Heaven, P. (2005) Personality and Preference for Academic Assessment: A Study with Australian University Students, *Learning and Individual Differences*, 15, 247-256
- Crossley, J., Humphris, G. & Jolly, B. (2002) Assessing Health Professionals, *Medical Education*, 36, 800-804
- Farley, J. (1989) The multiple-choice test: writing the question, *Nurse Educator* 14, 3-5.
- Fischer, M.R., Herrmann, S. & Kopp, V. (2005) Answering Multiple Choice Questions in High-Stakes Medical Examinations, *Medical Education*, 39, 890-894
- Fullerton, J.T., Parker, K.W. & Severino, R. (1997) Development and Outcomes of the Multiple-Choice Format National Certification Examination in Nursing and Midwifery, *Journal of Nurse-Midwifery*, 42, 349-354.
- Gardner-Medwin AR. (1995). Confidence assessment in the teaching of basic science. *Association for Learning Technology Journal*. 3, 80-85
- Hammond, E.J., McIndoe, A.K., Sansome, A.J. & Spargo, P.M. (1998) Multiple Choice Examinations: Adopting an Evidence-Based Approach to Exam Technique, *Anaesthesia*, 53, 1105-1108
- Holsgrove, G. (1992) Guide to post-graduate exams: multiple choice questions, *British Journal of Hospital Medicine*, 48, 757-761
- Lee, G. & Weerakoon, P. (2001) The Role of Computer-Aided Assessment in Health Professional Education: a Comparison of Student Performance in Computer-Based and Paper-and-Pen Multiple-Choice Tests, *Medical Teacher*, 23, 152-157
- Mccoubrie, P. (2004) Improving the Fairness of Multiple-Choice Questions: a Literature Review, *Medical Teacher*, 26, 709-

Race, P & Brown, S. (2001), *The Lecturers Toolkit* (second ed.), London: Kogan Page,

Schuwirth, L.W.T. & Van der Vleuten, C.P.M. (2004) Different Written Assessment Methods: What Can be Said About Their Strengths and Weaknesses?, *Medical Education*, 38, 974-979

Tarrant, M., Knierim, A., Hayes, S.K. & Ware, J. (2006) The frequency of item writing flaws in multiple-choice questions used in high stakes nursing assessments, *Nurse Education Today*, 26(8), 662-71