A Novel Enhancement to Multi-choice question assessment

Dr. Tim Hunt

Waikato Institute of Technology (Wintec)

`tim.hunt@wintec.ac.nz`

Rosanne S. Matheson and Derek Christie

Waikato Institute of Technology (Wintec)

Abstract

This paper describes a novel enhancement to the standard Multi-Choice Question (MCQ) type assessment. The new method utilises a time delay between students seeing the questions and when they are given the answers to choose from. During this period, students are encouraged to answer the question, as if they were attempting a constructed-response test. We argue that this modified test improves the ability of students to express their knowledge of the subject compared with a standard MCQ test. This is achieved while keeping the advantages of MCQ tests (e.g. efficient marking) that have made them a popular method of assessment. The details of how the enhanced MCQ assessment is to be delivered are explained, along with a description of the proposed qualitative and quantitative analysis of results. [1]

1 Introduction

Assessment practices are an important area of research because it is argued by Gibbs & Simpson, (2003, p. 22) that:

There is more leverage to improve teaching though changing aspects of assessment than there is in changing anything else and, at the same time, teachers know less about how students respond to assessment than anything else.

This paper describes for the first time a novel enhancement to multi-choice question (MCQ) assessment and discusses how it overcomes some of the drawbacks of this type of assessment and impacts positively on students' application of their knowledge in the multi-choice test.

In this paper we discuss a multi-choice test procedure which 'forces' students to work out the answer to the question without actually seeing the selection of answers to choose from and then asks them to select the answer that most closely agrees with their working.

We investigate whether the proposed enhancement encourages students to give greater thought to answering the questions than a standard MCQ test does and in the process increases the cognitive level that they employ during the test. The enhancement also reduces the level of inaccurate assessment, by reducing the opportunity for students to guess the answer rather than display the knowledge they have. This paper will be of interest to educators who want to use multi-choice assessment in a way which retains the benefits of this type of assessment while overcoming some of the drawbacks.

Our interest in the effect that multi choice tests can have on how students tackle a question and demonstrate their understanding grew from one of the author's computer programming courses. After the tests had been returned, one student asked for help, and in particular, she wanted to find out where she had gone wrong. Before, looking at the 'answer options' with her, the tutor attempted the question as if no 'answer options' were provided. Once the problem had been solved, the correct answer was obvious. The student believed that she too could have got the right answer if she had worked it out before looking at the 'answer options'.

2 Background

Increases in student numbers and reductions in academic staff have led to larger classes and heavier teaching workloads for many teachers in tertiary institutions. This means that teachers may lean towards MCQ assessment as it is time effective and efficient, offering ease of marking. This form of testing is also of growing interest to teachers utilizing e-learning in their courses as it can be easily administered and marked on-line. For these reasons it is likely that MCQ tests may be utilized more frequently by teachers as a form of assessment.

MCQ tests are also viewed favourably by students. Clarke, Heaney, Gatfield (2005), in research carried out with business students; found that MCQ tests are favoured by students because they do not disadvantage students with high intellectual and conceptual skills but poorer reading and writing skills. Kuechler and Simkin (2003) point out that as the emphasis on 'customer satisfaction' grows, students' assessment preferences may become an increasingly important consideration.

Given their important role in assessment, MCQ tests have been closely scrutinized and found to have a number of drawbacks. A considerable amount of research (see next section) has examined the perceived faults and many have come to the conclusion that more thought needs to go into the development of the questions. Fewer researchers have proposed that modification to the test itself will remedy one or other of the perceived faults.

---

3 Literature review

3.1 Benefits and drawbacks of multichoice assessment

The benefits of MCQs described in the literature (Ballantyne, 2002; Clegg and Cashin, 1986; Haladyna, 1997; Higgins and Tatham, 2003; Roberts, 2006; Williams , 2006) include their ability to offer objective and precise measurement of learning outcomes. Tests can be reliably marked as all answers are predetermined. Tests can be quickly marked by computer and can provide rapid feedback to students. This makes them especially efficient where large numbers of students are involved. Tests can be designed to assess the breadth of learning, and test a wide range of issues permitting a broad sampling of the content domain. Williams (2006, p.299) outlines the benefits of multi choice assessment for online learning and teaching and concludes that MCQ tests can also be used very effectively for formative purposes as an online, self-paced learning device.

There are also a number of perceived drawbacks of using MCQs and these have been extensively discussed in the literature (Ballantyne, 2002; Burton, 2001; Haladyna, 1997; Roberts, 2006) MCQ tests are said to be unreliable because of random guessing. A major task in using MCQ's is the creation of questions that are not easily and correctly answered by students who do not have a grasp on the subject being assessed. Students can become proficient at eliminating questions that are unlikely to be the correct answer, thus increasing the statistical odds of guessing successfully. MCQ tests do not assess a student's ability to develop and organise ideas and present these in a coherent piece of writing. It takes a long time to write plausible distractors - especially in cases where higher order cognitive skills are being tested. They fail to test critical or communicative skills and problem solving.

One of the most significant criticisms of MCQ assessment is that it can only test low level learning such as factual recall. Haladyna (1997, p. 36) states that 'much has been written on the underlying mental processes required in constructing versus selecting answers' and that this is a complex issue which requires further research. However like others *(*Clegg & Cashin, 1986; Higgins & Tatham, 2003; Killoran, 1992; Woodford and Bancroft, 2004*)* he believes that MCQs can test higher levels of student learning. A number of researchers have used Bloom's taxonomy (Bloom 1956) as a framework for designing questions across the 6 cognitive levels: knowledge, comprehension, application, analysis, synthesis, evaluation. Clegg and Cashin (1986) believe that multi-choice items can be written to evaluate higher levels of learning such as integrating material from several sources, critically evaluating data, contrasting and comparing information. William's(2006) research investigated the use of assertion-reason questions, a sophisticated form of MCQs that aim to encourage higher-order thinking on the part of the student. His findings suggested that ARQs were successful in generating reasoning rather than recall and are therefore an indicator of deeper learning but he still questioned whether students' performance may have had more to do with proficiency in English language (Williams, 2006, p. 291).

Another criticism is that MCQ tests encourage students to take a superficial approach to learning. There is growing interest in the idea that assessment measures tend to influence what is learned as well as the way in which it is learned. A growing body of research is focusing on the key features of assessment which promote learning. (Gibbs and Simpson, 2003; Angelo and Cross, 1993; Biggs, 2003; Black and William, 1988; Crooks, 1998; Loacker and Mentkowski, 2003; Zepke, 2003). This research suggests that the learning of students is very much driven by the assessments they undertake.

Paxton (2000) notes that there has been a move a way from MCQ testing because it is seen as limited and has negative effects on the quality of learning and teaching and the curriculum. Others (Haladyna, 1997; Taylor & Gardner, 1999) comment that when teachers and students expect their assessment to consist largely of multiple choice testing, 'multiple choice teaching' may be the result.

Despite the challenges, there is agreement that many of the potential problems with MCQ assessment can be 'designed out' with well written and constructed items.

> *'Many college teachers believe the myth that the multi-choice question is only a superficial exercise – a multiple guess – requiring little thought and less understanding from the students. It is true that many mc items are superficial but that is the result of poor test craftsmanship and not an inherent limitation of the item type.' (Clegg & Cashin. 1986).*

For this reason much of the attention given in the literature looks at the construction of the multi-choice question items. There has been less research that has focused on the design of the multi-choice test and the test procedure.

4 Methodology

4.1 Details of the course

The subject being assessed is an introductory level, computer programming development module (PD500 Program Development) on the DipICT(L5) course with approximately 40 students. The aim of the module is:

> *To provide students with a basic understanding of computer programming through the study of logic methods, software development concepts and documentation.* (ref 'New Zealand Institutes of Technology and Polytechnic Qualifications in Information and Communications Technology, Publisher NACCQ).

The students attend a weekly lecture that covers the 'theory' and then they have another 4 hours per week to put into practice what they have learned. Although the course concentrates on Program Development, the students are also being taught a practical programming

course during the same lecture and tutorial time slot. This gives them the immediate opportunity to turn their program development plans into actual programs.

4.2 The enhancement to the MCQ test

The enhancement proposed retains the advantages of the usual MCQ test and also combines the benefits of the MCQ and a constructed-response test, in which students are required to create their own answers rather than select the correct ones from a list of prewritten alternatives.

As discussed above, some of the learning outcomes of the course being assessed require students to work at the 'application' level of Bloom's (1956) taxonomy, demonstrating the ability to use learned material in new and concrete situations and applying rules, methods, concepts and principles.

As outlined in the literature review, it is in theory possible to construct MCQ tests that assess at all levels of Blooms taxonomy. However, it is acknowledged that this can be hard to achieve and so MCQ assessments may fall short of their intended level of assessment. Students may do well in an assessment by working at a comprehension level (they can recognize a correct answer) but would not have been able to independently construct the solution in a non MCQ question.

Our enhancement to the test procedure attempts to 'force' students to work out the answer to the questions without actually seeing the selection of answers to choose from and then asks them to select the answer that most closely agrees with their working. This can be achieved by using a 'time delay' between setting the question and giving out the 'answer options'. Further, if we limit the time that students have to select an answer, they are likely to have the greatest success by matching their workings (thinking at the application level) with one of the given solutions. It is proposed that this will reduce the ability to select the correct answer through comprehension alone or using 'test taking skills' thus negating some of the effects of sub optimal test questions.

It is useful to think of two distinct groups of students; competent and marginal. It is intended that this new enhancement will encourage the competent students to use their knowledge to successfully arrive at the correct answer – just as they would in a non MCQ test. They would not be confused by the 'distractors'. However, the marginal, students who can only work at the comprehension level or use 'test taking skills' will not have sufficient time to succeed using these methods alone.

4.2.1 The standard MCQ test (previous year)

Students were given questions each with a set of 'answer options' to select from. Once the student had decided which option was the correct answer to the problem, they used an electronic form (provided by the Moodle course management system (http://moodle.org/) to make their selection from a group of radio buttons. Paper copies of the questions and answers were also provided to aid students with reading the questions and in particular the diagrams.

To reduce the chances of students copying from each other, in the rather cramped computer room, the order of the questions were randomized in the Moodle environment. Students matched questions and answers based on the unique question name, rather than the order that the question appeared. This was an unwanted distraction for the students, but was deemed necessary to increase the difficulty of copying.

4.2.2 The enhanced MCQ test (this year)

The enhanced MCQ has a single point of difference in how it is administered. At the start of the assessment period, the students will be given the questions, but not the 'answer options' to choose from. The answer options will be given to the students one hour after the test has started. The students will then have 30 minutes to decide which answer option to select.

The students were told about the modified MCQ assessment that would be used well in advance of them taking the assessment. They were told that they would be expected to attempt the questions on paper during the first hour, although they would not be required to do so, as they would not contribute to the final mark. However, they were encouraged to make use of this time as it would be likely that they would then be able to make their selection, based on the answers they have constructed. The students will reminded about this at the start of the test.

All other aspects of the enhanced MCQ test are the same as the standard MCQ.

4.3 Proposed Analysis

Two methods of analysis will be used: 1) Statistical analysis of the marks gained, 2) Qualitative analysis of 'workings (doodles)'.

4.3.1 Statistical analysis of the marks

Assuming that this new test procedure is the only variable being changed, it is hypothesised that if the average score of the students' increases, then this increase is due to the new procedure. Student marks for two courses are to be collected for both 2005 (previous year) and 2006 (current year):

- The Program Development course (on which the enhanced MCQ will be used), referred to as PD,
- The 'control' course, Data Communications, referred to as DT.

The marks are to be entered into a spreadsheet in Data base form – one row per student with the year, a dummy variable for Innovation (0 if they were tested without the innovation and 1 if tested with the innovation) along with their DT and PD marks.

Note the 2005 and 2006 students are a completely different set of students, however student marks for PD and DT in a single year will only be used if the student did both PD and DT that single year (the students ID will

be used to ascertain this). Neither the student ID nor their name, or any other distinguishing attribute will be used for the rest of the analysis.

A regression will be done with PD as the response and Innovation and DT as predictors. A significant p value for Innovation will indicate a significant effect and the analysis will give an estimate of its size. The DT marks will act as a control between the years and the results for DT, while perhaps useful in other areas, have no meaning for us.

4.3.2 Qualitative analysis of workings

As all the students work will be collected, there is the opportunity to look at the answers the students wrote. This will give an opportunity for the researchers to assess if there is a high correlation between students who 'doodle' and gain a high mark. Being able to access the process that students go through in order to arrive at an answer may also suggest other significant factors relating to students' results and the design of MCQ tests similar to those found by Paxton (2000) in her study.

5 Further research

The authors would like to extend an invite to all conference delegates to participate in a multi-institutional study of this technique.

6 Acknowledgements

The authors would like to thank Wintec for permission to publish this work.

7 References

Angelo, T.A. and Cross, P.K. (1993): *Classroom assessment techniques*. 2nd Edition. San Francisco, Jossey-Bass.

Ballantyne, C & the Teaching and Learning Centre of Murdoch University (2002): *Multiple Choice Tests.*

Biggs, J.B. (2003): *Teaching for quality learning at University*. 2nd Edition. Buckingham, Open University Press.

Black, P. & Wiliam, D. (1998): Inside the Black Box: Raising Standards Through Classroom Assessment. *Phi Delta Kappan*, 80(2): 139-148.

Bloom, B. S. et al. (1956): *Taxonomy of educational objectives - the classification of educational goals: Handbook 1 Cognitive domain*. New York: David McKay.

Burton, R.F. (2001): Quantifying the effects of chance in multiple choice and true/false tests:question selection and guessing of answers. *Assessment and Evaluation in Higher Education,* 26(1):41-50.

Clarke, P., Heaney, J., Gatfield, T. (2005): *Multiple choice testing: a preferred assessment procedure that is fair to all our business students?* Retrieved on February, 16, 2006, from www.anzmac.org/ proceedings.

Clegg, V.L. and Cashin, W.E. (1986) *Improving Multiple-Choice Tests.* iDEA Paper No. 16, Retrieved on February, 16, 2006, from http://www.idea.ksu. edu/papers/Idea.

Crooks, T.J. (1988): The impact of classroom evaluation practices on students. *Review of Educational Research*, 58(4): 438-481.

Gibbs, G. & Simpson, C. (2003): *Does your assessment support your students' learning*. Retrieved June 20, 2005, from http://www.open.ac.uk/science/fdtl/ documents/lit-review.pdf.

Haladyna, T.M. (1997): *Writing Test Items to Evaluate Higher Order Thinking*. Needham Heights, MA, Allyn & Bacon.

Higgins, E. and Tatham, L. (2003): Exploring the potential of Multiple-Choice Questions. *Assessment. Learning and Teaching in Action,* 2 (1): Assessment, retrieved on February, 16, 2006, from http://www. ltu.mmu.ac.uk/ltia/issue4/higginstatham.shtml.

Killoran, J. (1992): In defence of multiple-choice questions. *Social Education*, 56(2):106-108.

Kuechler, W.L., & Simkin, M.G. (2003): How well do multiple choice tests evaluate student understanding in computer programme classes? *Journal of Information Systems Education,* 14(4):389-398.

Loacker, G. & Mentkowski, M. (1993): Creating a culture where assessment improves learning. in T. Banta (ed.), *Are we making a difference?* San Francisco: Jossey-Bass.

Paxton. M. (2000): A linguistic perspective on multiple choice questioning. *Assessment & Evaluation in Higher Education* 25 (2):109119.

Roberts, T.S. (2006): The use of multiple choice tests for formative and summative assessment, Australian Computing Education Conference, Hobart, Tasmania, Jan 2006. , retrieved on February, 16, 2006, from http://crpit.com/confpapers/CRPITV52Roberts.pdf.

Taylor, C. & Gardner, P.L. (1999): An alternative method of answering and scoring multiple choice tests. *Research in Science Education* 29(3):353-363.

Williams, J.B. (2006): Assertion-reason multiple-choice testing as a tool for deep learning: a qualitative analysis. *Assessment & Evaluation in Higher Education* 31(3):287-30.

Woodford, K. & Bancroft, P. (2004): Using multiple choice questions effectively in information technology education ASCILITE retrieved on February, 16, 2006, from http:/www.ascilite.org.au/conferences/pereth04/ procs/woodford.html.

Zepke, N. (2003): Reflecting-learning-teaching. In N. Zepke, D. Nugent and L. Leach, (eds.), *Reflection to transformation: A self-help book for teachers* (pp.17-33). Palmerston North: Dunmore Press.