Multi-choice Question Assessment with Time Delay

Tim Hunt
Waikato Institute of Tecnology
tim.hunt@waikato.ac.nz

Rosanne S. Matheson
Aviation and Travel Training Group

Derek Christie
Waikato Institute of Technology

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 are 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 (for example efficient marking) that have made them a popular method of assessment. Details of the first trial of this technique are given and although the main hypotheses were not proved by quantitative analysis, many qualitative benefits were observed.

Introduction

Assessment practices are an important area of research because it is argued by Gibbs and 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 work was first presented as a conference paper (Hunt, Matheson, & Christie, 2006). In the present paper we extend the literature review and document the results which previously had only been presented verbally.

The paper describes a novel enhancement to multi-choice question (MCQ) assessment and 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. The new enhancement '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 an incident that occurred in a computer programming course conducted by one of us. 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.

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 and 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 authors 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.

Benefits and drawbacks of multichoice assessment

The benefits of MCQs described in the literature (Ballantyne, 2002; Clegg & Cashin, 1986; Haladyna, 1997; Higgins & 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 MCQs 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 options 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.
- MCQs 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, Englehart, Furst, Hill, & Krathwohl, 1956) as a framework for designing questions across the six cognitive levels: knowledge, comprehension, application, analysis, synthesis, and 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. Williams (2006) 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 assertion-reason questions 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. (Angelo & Cross, 1993; Biggs, 2003; Black & William, 1988; Crooks, 1988; Gibbs, 2003; Loacker & Mentkowski, 1993; 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 away 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. According to Clegg and Cashin (1986),

> '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.

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.

Enhancing MCQ test procedures

Some research has looked at enhancing the test by requiring students to utilise additional material to answer the multi-choice questions. Other research devises ways where students show how they have arrived at their choice of answer. Higgins and Tatham (2003) found in their research with law students that by devising a strategy which involved students having to work through a set of additional materials to the MCQ test paper they were able to test the students' fulfilment of the 'higher level' learning outcomes (comprehension, application) of their course.

Woodford and Bancroft (2004) describe how MCQs can be used to test more than straight recall of facts and give examples which test students' comprehension of knowledge and ability to apply and analyse that knowledge. They suggest that sequentially dependent questions facilitate testing of higher cognition in IT education. Roberts (2006) adopted an innovative implementation of MCQ tests in which the primary aim was to enhance the process of learning and the assessing and grading of students was secondary. In his research, tests were sat and submitted on-line and multiple submissions could be made over a 24 hour period. A trial practice quiz was provided and students could refer to their notes and other resources as they sat the test.

Paxton (2000, pp.117-118) found in her research that MCQ tests could be powerful if ways could be devised to access the process that students go through in order to arrive at an answer. In her research with students studying Economics she found that students would have been able to answer questions more successfully had they been asked to express their understanding in their own words. She also found that the feedback on MCQ tests is often not very helpful because 'students are simply given a list of the correct answers and not shown the process by which the answer was arrived at'. She argues that 'think aloud protocols' on multiple choice questions might give us a sense of where the stumbling blocks occur for students.

The research done by Lister et al. (2004, p.133), which explores how students went about answering MCQs, is also of interest. They found that 'doodles' (where students were given paper on which they were allowed to draw pictures or perform calculations as part of answering the MCQs) and 'walk throughs' (a type of doodle where they systematically, manually execute (trace) a piece of code) make a

significant difference to students' success with MCQs: "if a student carefully traces though the code thus documenting the changes in variables, the likelihood of getting the correct answer is high".

Details of the course

The subject being assessed was an introductory level, computer programming development module (PD500 Program Development) on the DipICT (Level 5) 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.

The learning outcomes are:

1. Describe the program development cycle.
2. Apply the techniques of problem decomposition.
3. Apply a number of recognised logic depiction methods.
4. Desk check the methodology using test data.
5. Describe the purpose and characteristics of computer programming languages from first generation to current languages and reasons for their development.

The students attend a weekly lecture that covers the 'theory' and then they have another four 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.

Details of the assessment

The test is designed to assess the learning outcomes as listed above. Learning outcomes 2, 3 and 4 are the main outcomes of interest for this paper as they are concerned with demonstrating application of knowledge. The students are required to do the following:

- Draw/Select a structured diagram that meets the needs of a given problem description.
- Construct/Select a desk check, using given test data, that correctly checks a given structured diagram.
- Draw/Select a decision table that identifies all the conditions and actions of a given scenario.
- Draw/Select a decision tree that identifies all conditions and actions for a given problem.
- Draw/Select a logic table for a given Boolean operation, and evaluate a Boolean expression.

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 required students to work at the application level of Bloom's 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 greater 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.

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) 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 was 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.

The enhancement 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 were given the questions, but not the answer options to choose from. The answer options were given to the students one hour after the test had started. The students then had 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 their attempts 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 had constructed. The students were reminded about this at the start of the test. All other aspects of the enhanced MCQ test were the same as the standard MCQ.

Sample question

The following example shows a typical question that the students were required to answer:

Draw / Select Structure Diagram

For each person, the program prompts the user for the age and birth place of the person. If the age of the person is greater than 18, the program adds to a count of adults. If the birth place of the person is 'New Zealand' then, the program adds to another count of 'NewZealanders'.... Draw/Select a structure diagram to meet the above requirements.

After one hour, six diagrams were handed out for the students to choose from; two of which are shown in Figures 1a and 1b.

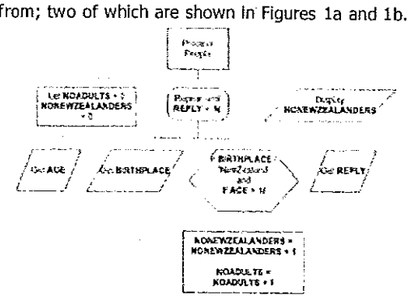

Figure 1a. An example of an answer option diagram that a student could choose.

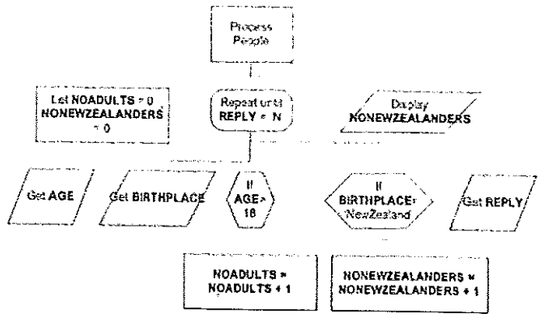

Figure 1b. An example of an answer option diagram that a student could choose.

Analysis

Two methods of analysis were used: statistical analysis of the marks gained, and analysis of 'workings (doodles)'.

Statistical analysis of the marks

Assuming that this new test procedure was the only variable being changed, it was hypothesised that if the average score of the students' increases with the innovation, then this increase is due to the new procedure. To allow for differences in student ability, a control group was required; however, due to ethical concerns about students being treated differently, instead of splitting the group into two, the marks for a second related course were used as a covariate to account for as much of these differences as possible. Student marks for the two courses were collected for both 2005 (previous year) and 2006 (current year):

- The Program Development course (on which the enhanced MCQ was used),
- The 'control' course, Data Organisation.

In addition, a dummy variable for Innovation was included - 0 if they were tested without the innovation and 1 if tested with the innovation.

Program Development (PD) and Data Organisation (DO) marks were only used for student who did both these subjects. (n = 66 of which 21 were in the Innovation group)

The basic model is -

PD mark = a + b x DO mark + c x Innovation

This assumes that the regression slope is the same with and without the Innovation term. The only difference is that the regression line for students who sat with the innovation (Innovation = 1) will be displaced a distance c from those without (Innovation = 0). The question is: can we be reasonably sure that the observed value of c is not due to random sampling?

The choice of which course to use, as the control course, was constrained by the available historic data for the group of students under study. The Data Organisation course was deemed to have a similar academic rigour requirement to Program Development and so would be suitable. Figure 2 shows the actual correlation observed.

Analysis of workings

As all the students' work was collected, there was the opportunity to look at the answers the students wrote. This gave an opportunity for the researchers to assess whether there was a correlation between the quantity of doodles and the marks achieved. 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.

Results

Hypothesis 1

If students are 'forced' to work out the answer before looking at the answer options they will achieve a higher mark.

Statistical analysis of the marks

The data was analysed using a full general linear model with interactions in Data Desk 6.0.1. The DO x Innovation interaction was insignificant (p = 0.84) so it is safe to assume that the regression slopes are the same with and without the Innovation term. When the data was reanalysed without the interaction (i.e. assuming equal regression slopes) we found that there was an average increase of about 2% on this data for those using the innovation. However, significance (p = 0.78). The low correlation between the Data Organisation and Program Development marks, while very real

($r = 0.41$, $p < 0.01$) means that the DO marks did not remove enough variation to make the innovation significant. It seems another means of allowing for student differences is required if the hypothesised effects have a chance of being seen.

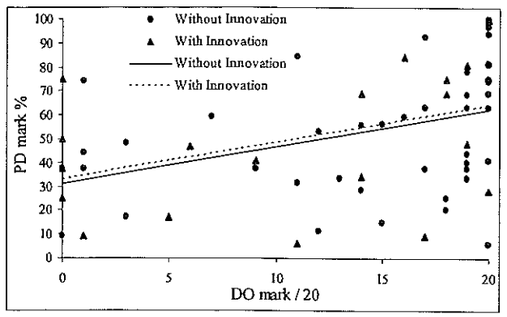

Figure 2. Program Development marks archieved by students with and without the innovation, assuming equal regression slopes.

Hypothesis 2

Students are required to work at the application level (Bloom's taxonomy) rather than the comprehension level.

It was expected that students working at the application level would reveal this by doing a large amount of 'working out' or doodles. The quantity of doodles was measured in two ways. The first method was to give a subjective value for quantity for each question. However, the subjectivity of this approach was a concern and so the second method looked to see if anything at all (yes/no) was written or drawn for each question and a total of (yes/no) doodles for each student was recorded. The two techniques were shown to correlate well to 'allow' the yes/no approach to be used as a measure of 'quantity of doodles'. Figure 3 shows a graph of the relationship between the mark that a student received versus the quantity of doodles that they made.

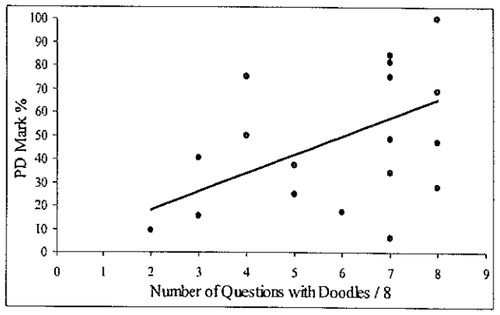

Figure 3. Program Development marks that students attained versus the quality of 'doodles' that they make

There was a weak but statistically significant correlation between the mark achieved and the quantity of doodles a student made ($r = 0.56$, $p = 0.01$). If the quantity of doodles is a good measure of working at the Application level then this implies that students who worked more at this level had a weak tendency to achieve better marks.

Hypothesis 3

If you limit the time that students have to look at the answer options they will be less likely to pass using 'test techniques'.

The time that students took to finish the test was automatically recorded by the Moodle test environment. Figure 4 shows the relationship between the time that students took to enter their answers and the marks that they achieved.

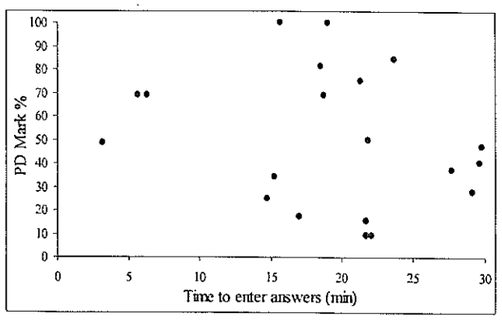

Figure 4. Relationship between how long a student took to enter their answers and the mark that they achieved

There was no significant correlation between a student's mark and the time taken to enter their answers ($p = 0.31$). As can been seen, most students finished before the 30 minutes given to them, so were not time constrained. Therefore the hypothesis was not tested. This also impacts on the third hypothesis: that 'marginal' students would attain a lower mark than with a normal multi-choice due to being time constrained. Clearly, 30 minutes was too long to constrain students and so even if the marginal group of students could be isolated, it is unlikely that an effect would be observed.

Discussion

Quality control

An often-cited drawback of multi choice questions is that students can guess the correct answer. The time-delayed technique allows the question setter to look at the 'doodles' to see if students are able to create the correct answer or are just guessing.

Learning experience

As well as being a method for assessing a student's ability, assessment should be an opportunity for students to learn. This can be hard to achieve with the standard multi choice technique as the assessor is not provided with any information on the thought process that the student was following. Likewise, even if the student is informed as to the correct answer, they are not usually given any feedback on what was wrong with their own thought process. Because the time-delayed technique encourages the student to work out their answers and the assessor is able to review those doodles, there exists the opportunity to turn the MCQ assessment into a learning experience rather than just an assessment. For example, the assessor could look at the doodles associated with questions that had a high percentage of incorrect answers and 'see' where the students went wrong. They could then use this as a tool for providing help/feedback to students after the assessment.

Students won't be tempted to learn/study for a multi-choice type assessment.

It has been reported (see literature review) that if students know that they are going to be assessed by multiple choice then they are likely to study for multiple choice questions. This work did not attempt to measure if students treated the assessment in this work as multi choice but as the students knew that their doodles were being collected, this may have affected how the students studied for the test.

Provides a tool to help assessors create new multi-choice questions

A major drawback with multi choice questions is the time and skill that is required to set suitable questions with corresponding plausible answers. The method described here provides a ready supply of answers that the students have written. The incorrect answers may be a source of good 'distractors'. Further, an opportunity is provided to remove questions that the students got right even though the 'doodles' were wrong or absent.

Inexpensive and easy to implement.

It is worth mentioning that the time-delay imposes virtually no extra cost or preparation over a standard multi choice assessment; especially if the time delay is administered manually. If an entirely automatic process is required, it is most likely that the software being used to administer the test would need to be modified.

Keeps benefits of standard multi-choice

The benefits outlined in this section are of course obtained while still keeping the main benefit (ease of marking) of the standard multi-choice question assessment.

Conclusion

This paper has described the use of incorporating a time-delay when administrating a multi-choice assessment. Although the major expected benefits (i.e. higher mark, working at application level, greater validity) were not conclusively demonstrated due to a low sample size, the assessment technique did produce a number of benefits that make the

technique useful regardless. These include an ability to review the quality of the multi-choice questions, being able to provide high quality feedback to students, and providing examples of possible 'distractors'. Where students are given multiple attempts to pass an assessment, this time-delay technique also encourages students to apply themselves during the assessment and so even if they do not pass on the first attempt, at least they have been through the thinking process, rather than just guessing, so providing a quality learning experience.

As stated previously, the students were told about the modified multi-choice assessment that would be used, well in advance of them taking the assessment. This in itself may change the level of preparation and learning that the students do for the test. How does this enhancement to the MCQ test impact on students' approaches to learning in the course? If they know that they will need to draw the diagrams, and not just select the correct one, students may modify their learning behaviour to suit the format of the assessment.

The ratio of 60 minutes for the first part of the test to 30 minutes for the second part of the test was conservatively chosen. The authors suspect that further gains may be made by moving this ratio towards 75/15 and so, if the arguments given in this paper hold, increasing the time that the students work at the application level. A suitable ratio needs to be found that still gives the students time to match their answer with the correct option.

Acknowledgements

The authors would like to thank Wintec (The Waikato Institute of Technology, http://www.wintec.ac.nz/) for permission to publish this work.

References

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

Ballantyne, C (2002). *Multiple choice tests.* Perth, Australia: The Teaching and Learning Centre of Murdoch University.

Biggs, J.B. (2003). *Teaching for quality learning at University.* (2nd ed.). 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., Englehart, M, Furst, E., Hill, W. & Krathwohl, D. (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 February 16, 2006, from http://www.anzmac.org/proceedings.

Clegg, V.L. & Cashin, W.E. (1986). *Improving multiple-choice tests.* iDEA Paper No. 16, Retrieved 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. & Tatham, L. (2003). Exploring the potential of multiple-

choice questions in assessment. *Learning and Teaching in Action, 2* (1). Retrieved February 16, 2006, from http://www. ltu.mmu.ac.uk/ltia/issue4/higginstatham.shtml.

Hunt, T.D., Matheson, R.S. & Christie, D. (2006). A novel enhancement to multi-choice question assessment. In S. Mann & N. Bridgeman (Eds.) *Proceedings of 19th Annual Conference of the National Advisory Committee on Computing Qualifications* (pp. 135-138). Wellington, New Zealand: NACCQ.

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.

Lister, R., Adams, E.S., Fitzgerald, S., Fone, W., Hamer, J., Lindholm, M., McCartney, R., Mostrum, J., Sanders, K., Seppala, O., Simon, B., & Thomas, L. (2004). A multi-national study of reading and tracing skills in novice programmers. *SIGCSE Bulletin, 36*(4), 119-150.

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), 109-119.

Roberts, T.S. (2006). The use of multiple choice tests for formative and summative assessment, *Proceedings of the Australian Computing Education Conference*. Retrieved 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. *Proceedings of the 21st ASCILITE Conference*. Retrieved February 16, 2006, from http://www.ascilite.org.au/conferences/perth04/procs/pdf/woodford.pdf.

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.