Multiple Choice Testing: A Preferred Assessment Procedure That Is Fair To All Our Business Students?

Peter David Clarke, Joo-Gim Heaney, Terry John Gatfield, Griffith University

Abstract

Multiple-choice questions are now a strongly preferred testing instrument across many higher education disciplines and business students in particular. This paper examines the reasons for their popularity and highlights the significant drivers that have fostered growth in the concept. Further, the paper takes an empirical study through the survey technique, to examine business students’ preference of multiple-choice tests against other testing approached. In addition, the paper explores whether multiple-choice tests show bias towards gender and nationality. It was found that there was no statistical difference on test results by gender and nationality but a difference in revision time between Australian and Asian students.

Keywords: assessment procedure, business students, teaching and learning, multiple choice

Background

One perspective of assessment considers the student to be an integrated component of the assessment selection and process. This approach requires a focus on the student to perceive, interpret, and integrate new information in order to form a coherent and meaningful whole that combines with their prior knowledge and former experiences (Dochy, Segers, and Buehl, 1999). It is evident that assessment practices have changed over time and there are increasing needs to consider and review new approaches as well as re-evaluate traditional assessment methods. As Birenbaum (1996) notes, the shift is from a 'test' or 'evaluation' culture to an 'assessment' culture.

It is now standard practice to include a mix of assessment items and approaches to test different aspects of the curriculum, to respect different learning styles and student preferences, and, to take advantage of technological development in education. New methods such as self and peer assessment, portfolio assessment and simulation methods have been developed and implemented in higher education situations. The role of stakeholders, such as the government, university policy makers, administrators, teachers and students are both directly and indirectly driving the assessment debate and practices that encourage and shape the call for efficiency and effectiveness in assessment approaches. Both the student and education providers experience the outcomes of this countervailing power

Outcomes for education providers results in larger classes and heavier teaching workloads that potentially lean towards assessment practices that are time effective and efficient. In the main, students generally fund their own education through government levies as well as participation in part-time or full-time employment. In addition, some students are not in close contact to be fruitfully involved with group work due to factors such as the urban sprawl, individual self-interests, and the increase in single supporting parents. Because of this singularity, students not only want to engage in good learning outcomes, but seek those assessments that involve minimal group project work, are relevant, low risk and need relatively limited test revision time. This has led to what Struyven et al (2002) have termed the ‘strategic’ or ‘achieving approach to learning’ in which students intend to achieve the
highest possible grades by using well-organised and conscientious study methods that pursue effective time-management processes (Entwistle, McCune and Walker, 2001).

Multiple choice testing

The concept of multiple-choice questions (MCQs) is not new (Bracey, 2000; Chan and Kennedy, 2002) and has been adopted in a variety of disciplines including physical education (Ayers, 2001), economics (Chan and Kennedy, 2002), mathematics (Beller and Gafni, 2000), psychology and fine arts (Delgado and Prieto, 2003) and the physical sciences (Chula, 2003). Ultimately, there is no one ideal assessment process and so too, MCQ testing has its advantages and disadvantage.

An aggregation of views (Ayers, 2001; Chan and Kennedy, 2002; Walstad and Saunders, 1998) suggests that advantages accrue to MCQs because the responses are easily scored with accuracy and objectivity; they provide fast feedback at relatively low costs and at frequent intervals. Instructors tend to favour recall questions because they are easy to design; however, questions can be set for memory recall, to apply knowledge or to evaluate given information. Importantly, the tests do not overtly disadvantage students with weak reading skills to the same degree as essay questions. Misreading one MCQ may loose the student a small percentage; however, misreading an essay question can result in substantial losses, hence, students with high intellectual and conceptual skills, but poor writing skills, are not disadvantaged with MCQs. Overall, MCQs can provide increased breadth and depth of coverage of material as opposed to essays that may concentrate on a focused area.

There are problems associated with MCQs as well (Ayers, 2001; Chan and Kennedy, 2002; Walstad and Saunders, 1998). Overall, multiple choice tests appear to be controversial and the major problems stem from poorly or inappropriately constructed test items (Paxton, 2000). It is recognized that creativity cannot be tested and this disadvantage reflects Bloom’s taxonomy of combining ideas to form a new whole (Bloom, Engelhart, Furst, Hill, and Krathwohl, 1956). However, the other five components of Blooms taxonomy: knowledge, comprehension, application, analysis and evaluation can be tested through MCQs.

The issue of MCQs in the area of higher order deep learning is therefore an on-going debate. Different types of assessment seem to encourage either a deep or surface approach to learning (Stryven, Dochy and Janssens, 2002). Zeidner (1987) found that multiple choice-type exams are generally perceived more favourably than essay-type items; they were considered to be low in students’ anxiety, perceived difficulty and complexity, while high in success, expectancy and feelings of ease. Therefore, it appears that MCQ tests are highly favoured by students on many different dimensions (Struyven et al, 2002).

In many instances, the literature strongly supports the fact that MCQs are able to provide information about student’s higher levels of understanding (Gronlund, 1998; Haladyna, 1994; Mehrens and Lehmann, 1991), rather than the notion that MCQ tests dumb-down higher order learning ideals (Higgins and Tatham, 2003). If measures are correctly designed the tests are able to assess the depth and breadth of students’ knowledge (Epstein, Lazarus, Calvano and Matthews, 2002). Consequently, as Yunker (1999) claims, MCQ tests can be authentic performance based measures. However, some students will engage in rote learning irrespective of the assessment method (Higgins and Tatham, 2003).
The research question in this study is whether there is indeed a preference by business students towards MCQ. There appears to be a high concentration of international students in business studies, this study also address the view that MCQ results are more favourable toward international students. Other aspects such as gender and length of study are also examined. The following hypotheses are presented:

**H1** There is a difference between student preferences and assessment types with a higher preference given to MCQ tests.

**H2** There is no statistical difference in the results of MCQ tests between male and female students.

**H3** There is no statistical difference in the results of MCQ tests based on student nationality

**H4** Asian international students study for longer periods for a MCQ test compared to Australian students.

**Methodology**

The issues of student assessment preferences and students’ engagement and performance in MCQ tests question were examined through a survey instrument comprised of a series of categorical closed questions, Likert scale questions and open-ended items. The questionnaire was administered to second year undergraduate International Marketing classes over two campuses at the commencement of a compulsory mid-semester test. The decision to conduct the research prior to commencement of the exam is appropriate because students are at a stage of application of learning and have stronger recall. Having post-exam participation means, with the relief of finishing the exam, students may misconstrue or not recall elements of their learning process. Staggered finishing time is also inherent in exams and thus generates issues relating to participation and disruption of others. This particular course was chosen because both classes were very large and they contained a very high percentage of international students. Apart from Australian and Asian student groups, there were four broad groups from the USA, Europe, Scandinavia and the Middle East. While students were not obliged to include their student number (this enabled their results to be compared with the survey responses), only 12% of the 593 students did not provide their student number and were eliminated from some of the statistical analyses.

Seven different items of assessment were identified as being relatively common to business students at the university level in Australia. These assessment item categories are short answers based on the text, short answers based on case studies, essay style long questions, semester long projects, take home final exams, as well as open-book tests and multiple choice tests; these formed the basis for preference ranking. In the MCQ tests students had to complete all items and negative scoring was not undertaken. Each item had a five-answer option. The test was constructed from a test bank supplied by the textbook publishers and tests of this type have been used successfully in prior years. All questions were selected carefully to ensure they were relevant to the student studying in an Australian context. The questions were relatively short and thus avoided unnecessary complexities because each question was based directly upon issues and statements found in the text. Additionally, the English grammar and structure was simple and clear, and reduced, as far as possible, any bias between domestic and international students.

The results showed the age range was between 17-60 with a mean of 21.6 and a standard deviation of 3.5. The data, as expected, was skewed to the right with about 4.4% of the
cohort over the age of 25, and the ratio between fulltime and part-time students was respectively 96% to 4%. With respect to the distribution of male to female, males comprised 47% of the sample and females 53%.

Examination of hypotheses

Australian, Asian, American and Middle East students all rated a first preference for MCQ tests while Europeans ranked MCQ test as a third choice. In a paired sampled t-test between all assessment item types it was found that at the .05 level, all preferences items were statistically different except for the between-items open book exams and the take home exams (p=.487). Hypothesis 1 was confirmed in that there is a substantive difference in student preferences and that MCQ tests are on average, the most preferred choices by students. The following results were ranked according to preference on a scale of 1 to 5 where 1 represented ‘strongly preferred’ and 5 indicated ‘definitely not preferred’.

Table 1: Student Assessment Preferences

<table>
<thead>
<tr>
<th>Assessment Item</th>
<th>Mean</th>
<th>Std Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple choice questions</td>
<td>1.95</td>
<td>0.96</td>
</tr>
<tr>
<td>Take home final exam</td>
<td>2.28</td>
<td>1.16</td>
</tr>
<tr>
<td>Open book tests</td>
<td>2.30</td>
<td>1.18</td>
</tr>
<tr>
<td>Short answer questions based on the text</td>
<td>2.48</td>
<td>1.08</td>
</tr>
<tr>
<td>Short answer question from case studies</td>
<td>2.99</td>
<td>1.12</td>
</tr>
<tr>
<td>Semester long projects</td>
<td>3.28</td>
<td>1.16</td>
</tr>
<tr>
<td>Essay style questions</td>
<td>3.58</td>
<td>1.11</td>
</tr>
</tbody>
</table>

Hypothesis 2 was tested through a t-test on gender and test score. The hypothesis was supported in that the test scores of male and females were not significantly different ($t_{(509)} = .001, p=.999$). For both males (n=241) and females (n=270) the scores were identical with a mean of 32.22/60 (54%) but with a standard deviation for males at 11.864 being slightly lower than that for the female group at 12.522. A one-way ANOVA was performed on nationality and test score. The results indicated there were no significant differences ($F_{(5, 241) = 1.573, p =.166}$). Thus, Hypothesis 3 related to student nationality was not confirmed; there is only a difference of 1.5 marks from a maximum score of 60 between the Australians (n=249) and the Asian group (n=270). A simple comparison of means was undertaken on length of time spent in revision for Australian and Asian students. For Australian students (n=237) the mean was 8.24 hours. This compares with Asian students (n=198) of 11.06 hours. Thus, the Hypothesis 4 is confirmed. Table 1 shows the distribution.

Table 1 Time Spent on Revision for MCQ Test

<table>
<thead>
<tr>
<th>Nationality</th>
<th>1-5 hrs</th>
<th>6-10hrs</th>
<th>11-15hrs</th>
<th>16-20hrs</th>
<th>&gt;21hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australian</td>
<td>82 (33%)</td>
<td>77 (31%)</td>
<td>59 (24%)</td>
<td>19 (8%)</td>
<td>10 (4%)</td>
</tr>
<tr>
<td>Asian</td>
<td>33 (13%)</td>
<td>59 (24%)</td>
<td>62 (25%)</td>
<td>44 (18%)</td>
<td>49 (20%)</td>
</tr>
</tbody>
</table>

One section of the questionnaire sought qualitative insights into students’ perception of MCQ and the first response item of open-ended responses were examined and coded into six categories. There was a similar response by Australian and the Asian students in that about 63% considered that ‘the answers were easy to guess’. This was followed by 18% of the
respondents saying ‘having the answers provided’ was important to them. The third category of ‘no difficulties with language and spelling’ related to about 8% of respondents, with the vast majority of these cases being Asian students. On the other hand, 73% stated that MCQs were confusing; 14% considered that they did not test their ability and 7% recorded that the questions were too many and tough.

Discussion and Implications

It appears that students of all nationalities prefer multiple choice testing assessment methods. Support for MCQ preference was provided in the qualitative part of the questionnaire by the use of positive statements such as being easy and having no difficulties with the language. The choice by students is no doubt influenced by many factors and some of these are an ease of course revision that lessens demand on student time and a reasonable assurance of low failure rate. The preference of MCQ testing may also be due to the relatively low failure rates of MCQ tests – students, by blind guessing on a five part test can provide on average, 20% of marks and even mediocre knowledge of the subject would improve on that score. However, this condition may be moderated by the degree of question difficulty set by the examiner for the test. In addition, for an essay question or a project, there is usually a significant call for a depth of research abilities, an understanding of integration of a topic and a need to exercise conceptual and creative skills. Hence risk of failure may be perceived to be higher in project style assessment than with MCQ tests.

Finally for international students, it is not necessary to have very advanced language skills to understand MCQ tests compared to other forms of assessment. It is normal in MCQ testing to restrict the questions to the set textbook with a defined genre, language and vocabulary used for the course of study. The statistical tests indicated that there was no difference between the groups in the areas of gender and nationality. This should assist in resting some of the contentious issues in the literature (see, for example, the discussion of Beller and Gafni, 2000). To find non-support for the hypothesis would have showed that MCQ tests were discriminatory and would normally be considered a questionably invalid instrument for use in higher education. There should be confidence by academics in administering MCQ tests based on such non-discriminatory findings. It is of interest that the Australian student, on average, spends at least 8 hours and the Asian student spends at least 11 hours in the study activity for a MCQ test. For the Asian student there is usually a handicap in overcoming English language difficulties and especially in the use of unusual disciplinary jargon that is sometimes coupled with complex western constructs.

Areas of future research

Irrespective of any perceived difficulties, this paper shows that students prefer MCQ; however, additional qualitative research needs to be undertaken to establish why there is strong support for MCQ testing. If it is mainly focused on the ‘easiest option’ without substantial learning objectives being satisfied, then, the concept must be re-evaluated. As was expected Asian students revised for longer periods than Australian students. However, objective measurements such as the diary method or peer corroboration are needed to substantiate how long students actually do study. With the increase in MCQ testing comes a call to examine any links between the testing approach and student-preferred learning styles. Furthermore, there is a need to undertake research through a controlled study to establish if deep learning does take place through MCQ testing. This last factor remains a highly
contentious issue. Finally, this paper’s research has been undertaken from the perspective of one business discipline and replication is called for across all disciplines.

Conclusions
This research demonstrates the popularity of MCQ testing held by students where there is no favour toward nationality or gender. Thus, on those factors MCQ tests should be considered to be a non-biased method of evaluating student understandings. This research supports student’s affirmation of the MCQ test instrument as an appropriate assessment format but it does not address the vexed issue of quality learning outcomes.
References


Stryven, K., Dochy, F. and Janssens, S., 2002, Student’s perceptions about assessment n higher education: a review, *Learning Communities and Assessment Cultures Conference*.
