Multiple-choice questionnaires and the success of law students: Results of a preliminary evaluation

Jeremy Prichard
University of Tasmania, Hobart, Australia
jeremy.prichard@utas.edu.au

Multiple-choice questionnaires (MCQ) are rarely employed in legal education in Australia. However, MCQ have a long history of use among US legal academics and, more generally, their use as low-weight, formative assessment is relatively non-contentious. This paper presents an evaluation of a low-weight MCQ introduced into a criminal law unit in 2010 at the Law School, University of Tasmania. The impact of this MCQ on students’ end of semester exam results were analysed with (a) data spanning 2007-11 and a total of 786 students, (b) student success on the MCQ and (c) a survey of students’ perspectives of the MCQ (N=137). The paper suggests there is some evidence that the MCQ has promoted student success. However, the paper also explores difficulties in conducting evaluations of assessment; mixed method designs are recommended.

Keywords: multiple-choice questionnaire, evaluation, law

Introduction

Considerable research has been conducted on the efficacy of psychometric testing. In academia psychometric assessment, in the form of multiple-choice questionnaires (MCQ), has been used across a variety of disciplines for many years (Bowland, Lester, & Williams, 2009). The use of MCQ among Australian legal academics appears to be relatively rare. By contrast, despite the fact that US legal academics continue to debate the use of MCQ, they are widely used in law school admission tests (Curcio, 2009). MCQ also “form the backbone” of most of the country’s bar exams (which are required for admittance into legal practise) (Case & Donahue, 2008, p. 372).

This paper presents the results of an evaluation into the use of MCQ as formative assessment in a criminal law unit at the Faculty of Law, University of Tasmania in 2010 and 2011. The first section of the paper discusses the literature concerning MCQ and their strengths and weaknesses as assessment pieces. The second section describes the rationale for introducing MCQ into the criminal law unit, as well as outlining key features of that unit regarding student numbers, teaching resources and so forth. In the third section data are presented which indicate that MCQ may promote student success. The paper makes recommendations for law teachers interested in (a) trialling MCQ as an assessment piece and (b) evaluating the impact of MCQ on student success.

Multiple-choice questionnaires (MCQ) in the US and Australia

Compared with Australia, in the US MCQ are a more dominant feature of the education landscape generally. Variants of the SAT, once an acronym for the Scholastic Aptitude Test, are used across the US as entry tests for college applicants. A very large body of exists on the efficacy of these tests (Cole & Gonyea, 2010), which have at times attracted intense criticism (Owen, 1985). MCQ are also employed in legal tertiary education in the US. As noted, MCQ are used: to
assess applicants for law colleges (e.g., Law School Admission Test); as means of assessment within degrees; and – by the legal profession – for bar exams (Curcio, 2009). This is not to suggest that MCQ are necessarily popular within the US legal academy. Some view MCQ as intellectually inferior to essay questions and too removed from the actual job of a lawyer (Case & Donahue, 2009; see further Varnava & Webb, 2009). However, while discussions continue among US legal academics as to how and when MCQ ought to be incorporated into assessment agendas, less prominent are voices querying if MCQ ought to be used in law (Daggett, 2007).

MCQ have a relatively low profile in Australian law schools. After conducting quantitative and qualitative research at 27 law schools nationally, Johnstone and Vignaendra (2003) produced a report on developments in legal education in Australia for the Australian Learning and Teaching Council, now known as the Office for Learning and Teaching. The report sought to, among other things, highlight best practice in teaching and learning and to map assessment methods. The report contains few references to MCQ – perhaps indicating the low level of interest among Australian legal academics in this type of assessment. An anonymous interviewee explained that their law school used “multiple-choice examinations” apparently in response to resource pressures (Johnstone & Vignaendra, 2003, p. 402). The statement did not clarify the pedagogical rationale for the MCQ, nor whether it was employed as formative or summative assessment. The only other relevant references in the document were interviewees’ opinions about the inappropriateness of MCQ in law. One interviewee suggested that MCQ were not appropriate for “many of the styles of learning and deep understanding that we want to achieve” (Johnstone & Vignaendra, 2003, p. 411). Another interviewee, a Dean, intimated that MCQ are the antithesis of best practice:

If you don’t have highly motivated, highly committed educationalists, what you end up with is people who would prefer to give a multiple-choice test than a 5,000 word research project ... That has the potential to do damage in learning outcomes (Johnstone & Vignaendra, 2003, p. 390).

As far as the research conducted for this paper was able to determine, since 2003 only one law-focussed article has been published concerning MCQ by Australian academics.

Some strengths and weaknesses of MCQ as assessment

MCQ design is complex, time consuming and hence potentially resource intensive. Even well designed MCQ can be unsuitable for assessing the depth of student learning (Biggs & Tang, 2007; Downing, 2005; Selby, Blazey, & Quilter, 2008) and can encourage surface learning as a study strategy (Case & Donahue, 2008). Standard exams at least assess communication skills (Biggs & Tang, 2007) and partially reflect legal-professional time pressures. Poorly designed MCQs can be ambiguous, illogical, too difficult (e.g. because of nuanced wording) or too simple (e.g. because answers can be deduced within or between questions) (Bowland, Lester & Williams, 2009).

The reputation of MCQs, for example as “multiple guess” tests (Owen, 1985, 33), may increase student anxiety prior to assessment (Faize, Dahar, & Niwaz, 2010). It may be difficult to alleviate such anxieties if teaching staff – because of their own lack of familiarity with MCQs – cannot advise students how to study for such assessment tasks (Fisher, 2008). Flawed MCQs can increase student failure rates. Downing’s (2005) study of flawed MCQs estimated that up to 15% of medical students were wrongly classified as having failed as a result of the flaws. Flawed or invalid MCQs are likely to create tensions between teaching staff and students (Fisher, 2008).
On the other hand, MCQs are certainly efficient to mark. Marking can be done electronically with software available at many universities. As Selby, Blazey and Quilter point out (2008, p. 204), this means that MCQ can circumvent or reduce the “emotionally and psychologically” draining aspects of high frequency marking for teaching staff. Students may appreciate the speed with which feedback is given via MCQ assessment. Teaching staff can be confident of greater consistency and objectivity in marking standards with MCQs.

MCQs also enable teachers to assess a greater breadth of student learning compared with other forms of assessment (Daggett, 2007); this reduces the chance that some students will perform better than others purely on the basis of the topics they chose to study (Case & Donahue, 2008). Case and Donahue (2008) point to empirical studies that have found a positive correlation between students’ scores on MCQs and more traditional forms of assessments, namely essays. This indicates that MCQs can integrate within existing assessment agendas to promote, among other things, Biggs and Tang’s (2007) concept of constructive alignment. While Biggs and Tang (2007) expressed concerns about MCQ as summative assessment, they acknowledge their potential usefulness as formative assessment pieces.

Proponents of the use of MCQ in legal education argue that most of the limitations of MCQs can be overcome with proper design, combined with a clear pedagogical rationale as to how MCQs can further intended learning outcomes (Case & Donahue, 2008; Fisher, 2008; Selby, Blazey, & Quilter, 2008). The science of MCQ design has been heavily influenced by the work of Haladyna (2004, as cited in Fisher, 2008; see overview by Bowland, Lester, & Williams, 2009). Fisher’s (2008) law-focussed guidelines for MCQ design, which are based on Haladyna’s (2004, as cited in Fisher, 2008) research, have arguably made quality MCQ design more achievable for legal academics without much experience in this form of assessment.

**MCQ in Criminal Law, University of Tasmania (UTAS)**

Criminal Law is a compulsory two-semester unit in the UTAS law degree. It averages between 140–170 students annually, with two lecturers and one tutor. The unit outline includes the following description:

> Criminal Law is a cornerstone of law students’ education. It involves studying the detail of particular offences – including murder, rape, assault and drug trafficking – as well as certain defences, such as ‘insanity’ and self-defence. Underlying these topics is an analysis of what the criminal law calls *mens rea*, or the guilty mind. Since the criminal laws studied are contained in Tasmanian and Australian legislation, the unit provides foundational skills in statutory interpretation, which are highly valued in various professional settings.

In Criminal Law students encounter complex legislative interpretation for the first time in their law degree. The unit is also intended to equip students with professional skills as future potential public prosecutors and criminal defence lawyers. In this sense it seeks to encourage functioning knowledge: pragmatic execution of legal principles and a legislative framework to novel hypothetical situations (Biggs & Tang, 2007). However, in Semester 1 especially, the course also involves declarative knowledge (Biggs & Tang) – knowing the structure of the main criminal law statutes and their interconnections.

In terms of the levels of understanding required in Criminal Law, five of Bloom’s (1956, as cited in Le Brun & Johnstone, 1994) hierarchical cognitive domains – namely, knowledge, comprehension, application, analyses, syntheses and evaluation – are employed in different dimensions of the unit. The intended learning outcomes for the unit include:
• recall and summarise legal definitions, the outcomes of court cases, and the facts that need to be proven to establish guilt for different criminal offences
• examine the reasoning behind a particular judgment, principle, proposition or interpretation
• appraise the social and international context of criminal laws and options for law reform
• select relevant legal principles and apply them in problem solving exercises involving hypothetical factual situations, and
• predict how a court might decide a question of law given a certain factual situation and conflicting legal opinion.

The unit’s assessment agenda pre-2010 is illustrated in Table 1, below.

<table>
<thead>
<tr>
<th>Table 1: Assessment agenda pre-2010</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Weight</strong></td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td><em>Legal research report</em></td>
</tr>
<tr>
<td>Tutorial participation</td>
</tr>
<tr>
<td>Tutorial paper x 2</td>
</tr>
<tr>
<td><strong>Sem 1 Exam</strong></td>
</tr>
<tr>
<td>Sem 2 Exam</td>
</tr>
</tbody>
</table>

*The legal research exercise forms part of a skills program to meet intended graduate attributes for the law course.

** A 5% MCQ was introduced in 2010 as formative feedback for the Sem 1 exam, which was reduced to 15%.

The content of the unit is integrated across both semesters. This means, for instance, that the summative Semester 2 exam can encompass concepts and material covered in the first weeks of the year.

**Rationale for introducing MCQ**
The Semester 1 short-answer 20% exam was introduced in 2007. The exam, occurring around June, was designed to operate as formative assessment. However, between 2007-09 the exam had high student failure rates (further discussed in Results, below). It was recognised that the assessment agenda did not provide students with feedback on intended learning outcomes 1 and 2.

In 2010 the unit coordinators introduced a further piece of assessment: a MCQ worth 5% of the year’s mark, conducted in Week 4 of Semester 1. The low-weight MCQ was designed to assess knowledge and comprehension of essential facts covered in the first weeks of the unit and so to work as formative assessment for the Semester 1 exam (the weighting of which was reduced to 15% of the year mark). The closed-book MCQ was designed to include over 30 questions, to
be conducted in-class (in standard lecture times and facilities) over 60 minutes. Most students completed the MCQ after 30 minutes.

**Validation process**

One lecturer designed the MCQ over approximately one day with reference to Fisher’s (2008) guidelines. The other Criminal Law lecturer then attempted to answer the MCQ, following the principal that ‘experts’ should agree on correct answers if the questions are free of flaws (Selby et al., 2008). This process identified some ambiguities in particular questions. Further changes were adopted after a tutor reviewed the MCQ. In 2010 and 2011 students’ perspectives of the MCQ were assessed in the end of year anonymous *Student Evaluations of Teaching and Learning (SETL)*. A Likert scale asked students to indicate the extent to which they agreed with the statement “The 5% multiple-choice test helped me to get a much better mark in the mid-year exam”.

**Feedback after the MCQ**

In both 2010 (36 questions) and 2011 (32 questions) feedback was provided in lectures on the MCQ. The correct answers were explained. In 2010 this process itself identified flaws in two questions, so all students were awarded an extra two marks (out of 36) regardless of the answers they had provided. Ten of the 2010 MCQ were repeated in 2011. Anecdotal reports in 2011 indicated that some students had benefitted from studying 2010 answers that they had sourced from the 2010 student cohort. This outcome is discussed further.

**Results from preliminary evaluation of the effectiveness of the MCQ**

Figure 1, below, presents the percentage grade distribution for the MCQ assessment in 2010 (N=135) and 2011 (N=174). Note that the grades were not adjusted.

![Figure 1: Student performance on MCQ 2010-11 (% grade distribution)](source: UTAS electronic data file)

<table>
<thead>
<tr>
<th></th>
<th>2010</th>
<th>2011</th>
</tr>
</thead>
<tbody>
<tr>
<td>NN</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>PP</td>
<td>27</td>
<td>22</td>
</tr>
<tr>
<td>CR</td>
<td>36</td>
<td>32</td>
</tr>
<tr>
<td>DN</td>
<td>24</td>
<td>19</td>
</tr>
<tr>
<td>HD</td>
<td>23</td>
<td>8</td>
</tr>
</tbody>
</table>

NN= fail, PP= pass, CR= credit, DN= distinction, HD= high distinction
Although the failure rate did not change between the years, clearly a greater percentage of students were awarded distinctions and higher distinctions in 2011. This result supports the anecdotal reports, noted above, that some students in 2011 had sourced the 2010 questions (10 of which were repeated in 2011).

As noted, the MCQ was introduced to provide early low-weight formative feedback for the Semester 1 short answer exam. Table 2, below, presents the percentage of students’ grades for the Semester 1 exam between 2007-11.

Table 2: Student performance on Semester 1 exam (% grade distribution) 2007-11

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td>NN</td>
<td>40</td>
<td>53</td>
<td>67</td>
<td>29</td>
<td>38</td>
</tr>
<tr>
<td>PP</td>
<td>23</td>
<td>18</td>
<td>16</td>
<td>27</td>
<td>31</td>
</tr>
<tr>
<td>CR</td>
<td>16</td>
<td>18</td>
<td>14</td>
<td>16</td>
<td>12</td>
</tr>
<tr>
<td>DN</td>
<td>13</td>
<td>8</td>
<td>1</td>
<td>19</td>
<td>13</td>
</tr>
<tr>
<td>HD</td>
<td>8</td>
<td>5</td>
<td>1</td>
<td>11</td>
<td>7</td>
</tr>
</tbody>
</table>

Source: UTAS electronic data file
Note percentages may not sum 100 due to rounding

Table 2, above, shows that the percentage of students who failed the Semester 1 exam rose from 40 to 67% between 2007 and 2009. Student numbers were consistent across this period, at about 160. In 2010, when the MCQ was introduced, the failure rate dropped to 29%, but this rose again to 38% the following year. The student cohort also fluctuated in these years; 135 in 2010 and 174 in 2011. Figure 2, below, presents the percentages of the grade distributions for the years, combining pass and credit (PP/CR) as well as distinction and higher distinction (DN/HD).

Figure 2: Student performance 2007-2011
With respect to high performance, Figure 2 above indicates that 21% of students in 2007 fell in the DN/HD bracket. A marked reduction in the DN/HD grades was recorded in 2008-09. This was followed by a sharp increase in 2010 (30%) and another drop in 2011 to 20%.

To attempt to measure the impact of the MCQ on the failure rate, the data were collapsed into two time periods: 2007-09 and 2010-11. The grades were also collapsed into two categories, namely fail (NN) and pass (encompassing PP-HD). See Table 3, below.

<table>
<thead>
<tr>
<th>Table 3: Semester 1 exam: fails and passes 2007-09 and 2010-11</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Fail</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Pass</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

Source: UTAS electronic data file

A chi-square analysis indicated that the difference between the two time periods was statistically significant, $\chi^2(1, N = 786) = 27.78$, $p < 0.01$, $\Phi = -0.19$. As noted, in the 2010 and 2011 SETLs (student evaluation questionnaires), students were asked to indicate the extent to which they agreed with the statement; *The 5% multiple-choice test helped me to get a much better mark in the mid-year exam.* In 2010 and 2011 just over 50% of students reported agreeing or strongly agreeing that the MCQ helped them to secure a “much better mark” in the Semester 1 exam. Those who disagreed or strongly disagreed varied between 20.5% (2010) and 15.2% (2011).

It is important to note that there was a low response rate for the SETLs in both 2010 (47%) and 2011 (43%). This appears to reflect the fact that many students prefer to listen to digital recordings of lectures rather than attend lectures and consequently they miss the opportunity to participate in the SETL process. All students are notified in advance when SETLs will be conducted in class, but this does not seem to improve response rates.

Discussion

There is some evidence that the MCQ achieved its goal in improving student success in the Semester 1 short answer, closed book exam. The failure rate for the Semester 1 exam dropped from 53.1% in the 2007-09 period to 33.9% in the 2010-11 period, when the MCQ was introduced. This difference was statistically significant. High performance – as measured by the number of students awarded distinctions or higher distinctions in the Semester 1 exam – also increased in the 2010-11 period. In terms of the students’ own perspective, SETLs conducted in 2010 and 2011 found that over 50% of students agreed (or strongly agreed) with the statement that the MCQ helped them to perform “much better” in the Semester 1 exam. This supports the view that the improvement in Semester 1 exam results found in the 2010-11 period is due, at least in part, to the introduction of the MCQ.

However, some important caveats must be noted. First, no attempt was made to control for the difficulty of the Semester 1 exam between 2007 and 2011. Although the same lecturer set the short-answer exam each year, fluctuations in its relative difficulty may have occurred that subsequently increased or decreased student success. Secondly, an experienced lecturer – who taught some of the Semester 1 content – retired at the end of 2007 and was replaced with an
early career academic (the author of this paper) in 2008. It is feasible that the students’ failure rate increased in 2008 and 2009 partly as the result of the teaching skills of the new staff member. The improvement in the 2010-11 period may be evidence of a learning effect in this staff member’s teaching skills.

As noted, fluctuations in the size of the student cohort deserve some consideration. Between 2007-09 the cohort averaged about 160. In 2010 the cohort was only 135. It could be suggested that improvements in student performance in 2010 were due to a lower staff to student ratio. However, Criminal Law offers 12 tutorials per fortnight. This means that each tutorial would average 13-14 students with a cohort of 160, and about 11 students with a cohort of 135. In these terms the change in student numbers does not appear important. In any case, staff to student ratio cannot explain the relatively strong student performance in 2011, which had a cohort of 174.

Some conclusions can be drawn about the design and implementation of MCQ assessment pieces. As noted, students were provided with full feedback on both MCQs, which is considered an important means through which formative assessment can assist learning (Biggs & Tang, 2007). Ten questions used in 2010 were repeated in 2011. This is probably the reason why students achieved higher grades in the MCQ in 2011 than in 2010. It may also explain why fewer students expressed dissatisfaction with the MCQ in the 2011 SETL.

However, the MCQ was designed to improve student success in the Semester 1 exam. Since student performance on this exam was stronger in 2010 than it was in 2011, it seems that the repetition of questions in 2011 diminished its effectiveness. Furthermore, repeating questions in 2011 may have unfairly advantaged students with established social networks, who would have been better able to source the 2010 feedback from friends. Developing a completely new MCQ may be the best system for ensuring fairness whilst being able to provide full feedback on the questions.

In conclusion, this paper suggests that low-weight, MCQs as formative assessment have potential to promote student success in law. However, legal academics considering introducing such assessment should not underestimate the time required to design and validate a MCQ. This paper illustrates some of the complexities involved in evaluating the effectiveness of a new piece of formative assessment. Where possible, evaluations should attempt to control for the difficulty of the original assessment agenda. A lesson learned from the evaluation presented in this paper is that although quantitative analyses are valuable, these should be matched with qualitative research techniques, particularly in relation to students’ perspectives.

References


Copyright © 2012 Jeremy Prichard. The authors assign to HERDSA and educational non-profit institutions a non-exclusive license to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The authors also grant a non-exclusive license to HERDSA to publish this document in full on the World Wide Web (prime site and mirrors) and within the portable electronic format HERDSA 2012 conference proceedings. Any other usage is prohibited without the express permission of the authors.