A Teaching Practice Review Of The Use Of Multiple-Choice Questions For Formative And Summative Assessment Of Student Work On Advanced Undergraduate And Postgraduate Modules In Engineering

Aidan O'Dwyer†

†School of Electrical Engineering Systems
Dublin Institute of Technology

Abstract

This paper reflects on, reports on and evaluates the use of multiple-choice questions, in both formative and summative assessment modes, on control engineering modules taken by advanced undergraduate and postgraduate engineering students, over four academic years. The material covered by the modules is analytical in nature. The rationale for examining the assessment strategy was the poor performance by students in a closed book traditional terminal examination on one of the advanced undergraduate modules over some years. The change in assessment strategy has improved student learning, as measured by assessment data; student feedback on the assessment methods is broadly positive. The assessment approach has been extended with success to some postgraduate modules. Overall, an evidence based approach is taken in the reporting and analysis.

Keywords: assessment, multiple-choice questions, engineering.

1. Introduction

The use of multiple-choice questions in assessment has been reported extensively in the engineering education literature and the wider educational context; for example, excellent online guides on the topic are available (e.g. Multiple Choice and Matching Tests, 2005). In the Irish context, disciplines that use multiple-choice tests include engineering (e.g. IIE Ireland diploma in industrial engineering programme www.iie.ie/education.asp), medicine (e.g. membership examination of the Royal College of Surgeons http://www.intercollegiatemrcs.org.uk/old/examination_html), business (e.g. Insurance Institute of Ireland qualifications http://www.iii.ie/upload/courses_and_credentials/files/1301562545_MCQ_Study_Guide_2011_30th.pdf) and language learning (e.g. Teastas Eorpach na Gaeilge http://www.teg.ie/english/info_advice_candidates.htm).

Multiple-choice questions can have two choices of answers (true/false), though, more commonly, four choices of answers are available. It is recognised that raw scores from these tests should not be used directly. The reason is that, for example, in a test with a passing threshold of 40%, where each question has four choices of answer, a student may know the answers for 20% of the questions and guess the answers correctly for one quarter of the rest of the questions, passing the examination. Scaling may be done using a probabilistic approach (Zhao, 2005, 2006) or a simpler approach (which employs negative marking). The scientifically sound probabilistic approach suggests that the optimum number of choice of answers for questions is 4. In addition, if the number of questions is greater than 18, for example, there is less than 1% probability of obtaining a scaled mark of 40% by pure guesswork. This probability falls to less than 0.01% if the number of questions set is greater than 48 (Zhao, 2005, 2006).

There is a vigorous debate in the literature about the role of multiple-choice questions in assessment. The advantages suggested for the use of such questions fall into three categories:

- They facilitate comprehensive student learning. According to Excell (2000), the questions have the potential to cover the whole of the syllabus and they force the student, in principle, to learn all the taught material, though it is also suggested that multiple-choice questions are not suitable for assessing numerical design exercises, a feature of many traditional engineering examination questions (Excell, 2000; Zhao, 2005). In a discipline-specific comment, Fenna (2004) suggests that multiple-choice tests are particularly desirable in engineering, to require the student to learn and correctly apply fundamental knowledge. In addition, multiple-choice tests are effective self-assessment tools (Davies, 1994; Azalov et al., 2004; McElroy et al., 2006). It is suggested that multiple-choice tests are especially suitable for knowledge-based subjects that are well defined, do not change rapidly with time and have unambiguous right answers (Excell, 2000; Azalov et al., 2004; Zhao, 2005).
• Assessment efficiency. Well designed multiple-choice tests are an efficient means for the assessment of knowledge, analytical ability, language proficiency and numerical skills involving a large number of examinees (Zhao, 2005) and the tests are suitable where the relative competence of the examinees in a large sample size is to be assessed (Zhao, 2005). Automatic marking is possible (Excell, 2000) and results can be obtained quickly (Brown, 2001; Zhao, 2006), with test scores being reliable (Chang et al., 2007).

• They are amenable to analysis. Among the tools available are numerical measures of the quality of the individual multiple-choice question based on student selection of the answers and measurement of the ability of the question to discriminate between capable and weak students (Brown, 2001). A variety of other metrics, such as the Kuder Richardson 21 (KR-21) reliability measure and the standard error of the measurement (SEM) of the test, may be readily employed (Ebel and Frisbee, 1991).

However, there are objections to the use of multiple-choice questions in assessment, which tend to fall into two broad categories:

• General concerns about suitability. It is suggested, for example, that multiple-choice questions tend to address superficial facts, which may encourage learning of surface detail (O’Loughlin and Osterlind, 2007), though Struyven et al. (2006), among others, do not agree. O’Loughlin and Osterlind (2007) also suggest that even if the question is carefully worded, assessors cannot be sure that a student who answers correctly not only knows the correct answer but also understands the subject being examined and that students can select a correct answer for superficial reasons, such as selecting the answer through a process of elimination. However, it is possible to test student cognitive skills with properly constructed multiple-choice questions (Azer, 2003; DiBattista et al., 2004; Chang et al., 2007). It is also suggested that students find multiple-choice questions ‘confronting’ and would prefer to express themselves more fully (Brown, 2001). Perhaps the most extreme such comment is that of Azalov et al. (2004), who suggest that multiple-choice tests trivialise educational measurement. Separately, there are concerns about suitability for assessing particular learning outcomes e.g. numerical design exercises, as mentioned already.

• Structural issues. One concern is that students may select answers at random, though negative marking or other strategies can be used to reduce this (Excell, 2000). Whether to employ such correction is a vexed issue, discussed comprehensively by Betts et al. (2009). It is suggested that multiple-choice questions are ‘very much more difficult’ to write than descriptive questions (Brown, 2001). In addition, the number of possible questions rapidly becomes rather limited (Excell, 2000). However, textbooks that contain large numbers of multiple-choice questions are available in engineering (e.g. Floyd, 2007), optometry (e.g. Fletcher and Oliver, 1996), physiology (e.g. Colbert, 1996), mathematics (e.g. Bolt and Reynolds, 1978) and physics (e.g. Porter, 1987), among other disciplines.

These advantages and disadvantages suggest that multiple-choice questions should be used as one strand in a balanced and creative summative assessment regime, matched to the learning outcomes that are being assessed.
2. Assessment Outline

Multiple-choice questions are used in formative and/or summative assessment mode by the author in a number of advanced undergraduate (Level 8) and postgraduate (Level 9) programmes and modules. The details are summarised in Table 1.

<table>
<thead>
<tr>
<th>Programme</th>
<th>Level and stage</th>
<th>Module name</th>
<th>Year and semester</th>
<th>Formative and/or summative</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>DT235 BSc in Medical Physics and Bioengineering</td>
<td>Level 8, Years 2 and 3</td>
<td>Feedback and Control</td>
<td>2009-10, S1</td>
<td>Both</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2010-11, S1</td>
<td>Both</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2011-12, S1</td>
<td>Both</td>
<td>16</td>
</tr>
<tr>
<td>DT021 BE in Electrical/Electronic Engineering</td>
<td>Level 8, Year 4</td>
<td>Control Engineering</td>
<td>2008-9, S1</td>
<td>Summative</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2009-10, S1</td>
<td>Both</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2010-11, S1</td>
<td>Both</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2011-12, S1</td>
<td>Both</td>
<td>20</td>
</tr>
<tr>
<td>DT015/DT711 MSc in Energy Management</td>
<td>Level 9</td>
<td>Delay Systems</td>
<td>2008-9, S1</td>
<td>Summative</td>
<td>5</td>
</tr>
<tr>
<td>DT702/DT703 ME in Pharmaceutical Process Control and Automation</td>
<td>Level 9</td>
<td>Energy Control Systems</td>
<td>2008-9, S1</td>
<td>Summative</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2010-11, S1</td>
<td>Both</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2010-11, S2</td>
<td>Both</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2011-12, S2</td>
<td>Both</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Process Control</td>
<td>2009-10, S1</td>
<td>Both</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2010-11, S1</td>
<td>Both</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Advanced Control</td>
<td>2009-10, S2</td>
<td>Both</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2010-11, S2</td>
<td>Both</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 1: Details of Level 8 and Level 9 programmes and modules which use multiple-choice questions (S1, S2 = Semester 1, 2; n = number of students).

When formative assessment is used, students submit answers to short, paper-based, multiple-choice quizzes, attempted in the classroom; for Level 8 students, each quiz had a mean of six multiple-choice questions, with four answer choices. The quizzes are typically given at the end of each topic explored in the classroom. Students complete the quizzes anonymously, with the author communicating that the purpose of the work is to identify “muddy points” in student understanding. The solutions of the quizzes are then explored in the next classroom session, before a new topic is started. This allows the multiple-choice quizzes to guide learning and teaching, and is compatible with work done by Nicol (2007), for example, in linking multiple-choice testing to the principles of good feedback.

Multiple-choice questions are used as part of the summative assessment regime linked to particular module learning outcomes. This is detailed more fully in the following sections which treat the Level 8, Year 4 Control Engineering module, and the Level 9 modules summarised in Table 1, respectively. In all summative assessments with multiple-choice questions, negative marking is used if the solution offered is incorrect.
3. Level 8, Year 4 Control Engineering Module

The module has 5 European Credit Transfer and accumulation System (ECTS) credits and is scheduled for the thirteen-week first semester of the final year of the BE in Electrical/Electronic Engineering. It is the third of a series of four linked modules, over four semesters. The Control Engineering module stream, which consists of the final three of the four linked modules, is chosen by students at the end of second year, after the completion of the first of the four modules. This stream is one of five specialist module streams on the programme. Students have a restricted choice of specialist module streams, and typically, 30% of the students on the programme choose the Control Engineering module stream. For each of the linked modules, the subject is composed of two portions, a more analytical portion labelled Control and a more applied portion labelled Automation. The author has responsibility for learning and assessment for the Control portion of the third Control Engineering module mentioned. The learning outcomes of the Control portion of the module are that the learner will be able to:

1. Demonstrate the ability to synthesise fundamental modelling and controller design concepts.
2. Effectively present an untaught control applications topic to their peers, and in turn assess other presentations in an ethical manner;
3. Design and analyse controller solutions, with the assistance of an appropriate interactive tool;
4. Display knowledge in system modelling and controller design, using a state-space approach.

These learning outcomes contribute to the six programme outcomes required for Level 8 programme accreditation by Engineers Ireland (2007); three programme outcomes are technically oriented, with the remainder concentrating on the social, business and ethical context. Learning outcomes 2 and 3 address, to varying extents, the full range of programme outcomes, and are assessed continuously; such learning outcomes are less suitable for assessment using multiple-choice questions. In the 2011-12 academic year, the assessment regime for learning outcomes 2 and 3 is, respectively:

- A team-based PowerPoint presentation on a control engineering topic (in Week 6). This is peer and tutor assessed with the aid of a rubric, and is worth 20% of the module portion mark.
- An open-book design exercise, with the assistance of a computer based interactive learning facility (in Week 9). This is assessed by the tutor and is worth 25% of the module portion mark. A sample solution was provided to the students, with a marking rubric.

Learning outcomes 1 and 4 address mainly the first three Engineers Ireland programme outcomes, and it has been traditionally difficult to ensure that learners fulfil these learning outcomes, as they are analytical and wide ranging; one of the Level 8 programme external examiners for 2011-12 wrote, in this context, that "sometimes a course on control can be theoretically demanding and yet at the end of it, many students will display little real
understanding of the subject”. The author has become familiar with this problem over many years; for example, in the last traditional terminal closed book examination in the previous accredited version of this module, in the 2007-8 academic year, student performance in Control (mean=25%, standard deviation=21%) was much lower than in Automation (mean=58%, standard deviation=24%); overall, for the module, mean=42%, standard deviation=18%, n=17. The mean percentage was considerably less than that scored by students in comparable modules.

In this context, the author decided to assess learning outcomes 1 and 4 using open-book multiple-choice quizzes, to encourage student engagement with the material, using “the fairest way of testing knowledge of a wide-ranging syllabus” (Heywood, 2000). The open-book assessment approach, it was felt, eased pressure on students, bearing in mind the assessment timescale. The learning outcomes are assessed by:

- Two 50-minute open-book multiple-choice examinations (Weeks 2 and 3). Each examination consisted of 25 questions, with 4 choices of answers. These assessments are weighted at 25% of the module portion mark.

- Two 50-minute open-book multiple-choice examinations (Week 13). Each examination consisted of 25 questions, with 4 choices of answers. These assessments are weighted at 30% of the module portion mark.

In addition (taking the module in 2011-12 as an example), seven paper-based multiple choice quizzes were used for formative assessment and feedback to students throughout the module. Quizzes had from 4 to 8 multiple-choice questions (44 questions in total).

The purpose of the multiple-choice examination in Weeks 2 and 3 is to assess the pre-requisite knowledge for the module, as recommended by Felder and Brent (2001), among others. This was considered necessary, as students had successfully completed the previous module in the subject nine months previously (with the intervening time devoted to a full-time work placement). One advantage of using multiple-choice questions for assessing pre-requisite knowledge is that the instructor can, by analysing the percentage of students correctly answering each individual question, assess the topics that students find most difficult, and treat these topics subsequently. This was done by the author.

The open-book examinations in Week 13 were moderated by an external examiner. Examiner feedback is encouraging. The external examiner in 2010-11 remarked: “multiple choice examinations in year 4 have proved controversial here ... but it is obvious that a lot of useful material/learning outcomes can be achieved with such an examination. As is always the case I would have slight concerns in relation to the fairness of negative marking in this context ... I have complete faith in the efficacy of the exam as has been laid out”. A new external examiner in 2011-12 remarked: “I very much like the underlying philosophy of this course, in the context of the availability of other more traditional control courses in your degree programme. I do understand the importance of testing the students' background in week 2, to see where they are starting from – this helps in targeting the course. I have had a good read through the open-book examination .... in general, it is well written, though I have enclosed some very minor suggestions .... I believe that the questions are well thought out and are relatively
straightforward. This paper has been well designed to determine whether the students understand the basic concepts of control. Good students, who understand these concepts, will score very highly in this paper. Since this is a multiple choice paper, the questions are naturally not very involved from a mathematical or problem solving point of view. I suppose that the problem solving aspect of control design will be dealt with in the design exercise component ...”.

The open book multiple-choice examination structure early and late in the module has remained consistent over four academic years. When student assessment data is analysed for 2008-12 inclusive, a number of conclusions emerge:

- There is a highly statistically significant, positive correlation between performance in the open-book multiple-choice examinations and performance in the other assessments (n=59, p<0.00005, r=0.51). The correlation figure of 0.51 appears well balanced between an unduly low and high correlation figure. An unduly low correlation figure would suggest that the multiple-choice examinations are assessing rote learning of factual material, suiting students adopting a surface learning approach, assuming, as is likely, that the presentation and design exercise is assessing higher-order skills linked to deep learning. On the other hand, an unduly high correlation figure would suggest that the different types of assessment are assessing the same learning outcomes, which is undesirable.

- There is a highly statistically significant, positive correlation between performance in the open-book multiple-choice examinations taken early in the module and those taken late in the module (n=58, p<0.00005, r=0.55). In this case, the moderate correlation figure suggests that the two sets of multiple-choice examinations do not have redundancy i.e. they are assessing different learning outcomes.

- There is a highly statistically significant, positive correlation between the overall student performance in the subject and performance in Automation, the applied part of the module which is also completely continuously assessed (n=59, p<0.00005, r=0.54). Student performance in Control (mean=50%, standard deviation=12%, n=59) is still lower than in Automation (mean=65%, standard deviation=15%, n=59), but the difference is much less than in 2007-8, for example (referred to previously), indicating better fulfilment of the learning outcomes of the Control section of the module.

These results suggest that assessment using multiple-choice questions is valid and allows students greater facility to demonstrate their knowledge of the subject, when compared to a closed book traditional terminal examination.

The author also obtained student feedback on their learning experiences in the Control section of the module, including the use of multiple-choice tests, through a standard student survey questionnaire, discussion with the class representatives, and, in 2009-10, from an informal student focus group. Space does not permit student feedback to be recorded fully. In summary, in the first academic year this learning took place (2008-9), students were critical of their learning experiences, writing in a formal letter to the head of department that “we were informed beforehand that the structure of this assessment was different ... and that we were the first year to take this type of assessment. However, we believe it is grossly unfair as we are
losing marks as a result of this method of examination compared to someone who took the control exam last year, in its more traditional format. We understand that exam formats have to change and that the new format must be tried and tested with a class group. But we are all agreed that using a final year as a test group is completely unfair”. This feedback motivated the author in subsequent academic years to discuss, in the first lecture of the module, learning outcomes for the module and the manner in which the module would be assessed. Then, the author discussed the continuous learning approach to be adopted in the module, stating that module content builds on previous work and reminding the students of the learning outcomes for the previous two control modules. Chapters 2 to 19 of Wilkie et al. (2002) were referenced, with the author suggesting that the multiple-choice questions would be similar to a selection of those at the end of each chapter in this book; a sample multiple-choice examination, with solutions, was also provided on the college virtual learning environment. The author also increased the time students were permitted to spend answering the multiple-choice summative assessments, and introduced formative multiple-choice quizzes throughout the module. In subsequent student feedback though the informal focus group, these quizzes were generally considered to be helpful by students. The open-book nature of the assessment was generally considered fair, though concern was expressed about negative marking in the summative assessments. In subsequent discussion, a number of students suggested that choosing the correct answer, with a penalty for choosing the wrong answer, was too demanding. Interestingly, when the author suggested that it was ethically important for engineering professionals not to propose solutions to problems in which they had less than full confidence, two students disagreed based on their work placement experience. They suggested that accuracy was not demanded of them, as their work was subsequently checked for errors. This issue provided a theme for fruitful discussion in the first lecture of the module over the past two academic years.

In these two academic years (2010-11 and 2011-12), student concern, as communicated through the student questionnaire, focused more on perceived shortcomings of peer assessment of the PowerPoint presentation rather than on issues associated with the multiple-choice examinations. There was still some concern expressed about the multiple choice assessment of pre-requisite knowledge; a typical student comment on this issue is “25% of module mark on a review assessment is very high”. On the other hand, the formative multiple-choice quizzes were still considered helpful, a typical comment being that “regular anonymous MCQ tests were very effective practice and showed what the class didn’t understand”.

Table 1 also reveals that multiple choice questions are used, in formative and summative assessment mode, for learning and assessment on other Level 8 modules. When used in the terminal examination, the question with multiple-choice parts is either compulsory (Delay Systems) or optional (Feedback and Control), depending on the module learning outcomes. Space does not permit a full discussion; broadly, teaching and learning experiences are similar to those explored in detail for the Year 4 Control Engineering module.
4. Level 9 Modules

In brief, for the summative assessment of the Level 9 modules listed in Table 1, one question is made up of multiple-choice parts in the closed-book terminal examination; this question is either compulsory (Energy Control Systems, Process Control Engineering) or elective (Advanced Control Engineering), depending on the module learning outcomes. This discussion will concentrate on the learning and assessment of the 13-week duration, 5-ECTS credits Energy Control Systems module, taken by 62 students, from a wide variety of both engineering and non-engineering backgrounds, since the 2008-9 academic year. Learning on this module concentrates on the technical, business and strategic case for using automatic control to improve energy efficiency. The assessment regime (in 2010-11 and 2011-12) was as follows:

- A team-based PowerPoint presentation on an environmental feedback system topic (in Week 8). This is peer and tutor assessed with the aid of a rubric, and is worth 25% of the module mark.

- An individual open-book multiple-choice assessment (in Week 12). This consisted of 50 questions, with 4 choices of answers. This is assessed by the tutor and is worth 25% of the module mark. Generally, students find the assessment challenging, with performance mean and standard deviation being 46% and 18% respectively (n=48). Student feedback reflected this, with a typical comment in 2010-11 being “Assignment 2 (MCQ quiz) was hard for me – technical language, need to answer all the questions, negative marking”.

- A closed-book terminal examination after the completion of the module. Consisting of four questions, the compulsory first question was in multiple choice form with 24 parts, with 4 choices of answer for each part. The other three questions were conventional examination questions, with learners required to attempt two such questions. The closed-loop terminal examination was monitored by an external examiner, with the external examiner in 2011-12 remarking that “the marking system for question 1 is ingenious and generally is appropriate”.

In addition (taking the module in 2011-12 as an example), nine paper-based multiple choice quizzes were used for formative assessment and feedback to students throughout the module. Quizzes had from 4 to 13 multiple-choice questions (77 questions in total). As with the Level 8 module, learners tended to find these quizzes beneficial for learning, with a typical student feedback comment being “MCQs throughout course were very helpful”.

5. Discussion And Conclusions

In this paper, the author has attempted to undertake a systematic review of his teaching practice, with an emphasis on the evaluation of, and reflection on, the use of multiple-choice questions in both summative and formative assessments. Attention has been drawn to the extensive literature on the use of multiple-choice questions in assessment; indeed, Burton (2005) suggests wryly that “it includes too many textbooks and journal articles for most of us to read”.

In conclusion, multiple-choice questions, in both formative and summative assessment mode, have the advantages of facilitating comprehensive student learning, and allowing analysis of student understanding. The evidence is that the multiple-choice assessments are well matched to the learning outcomes with which they are associated, with other learning outcomes assessed using other appropriate approaches. The small number of students in the author’s modules means that assessment efficiency is not a priority; it also means that it is difficult, in some cases, to show statistically significant relationships from the assessment data. The author suggests that many Level 8 and Level 9 programmes in engineering have learning outcomes that make multiple-choice questions suitable as an assessment strategy (as part of a suite of assessment options). Students are more favourable to the use of multiple-choice questions in formative assessment, with negative marking of these questions in summative assessment attracting adverse comment.

The author intends to continue to explore the use of multiple-choice questions in the modules for which he has academic responsibility. The author is presently considering placing a bank of multiple-choice questions on a virtual learning environment, allowing enhanced student access to formative assessment and learning. There is also potential in using a standard multiple-choice “concept inventory” test to examine student understanding of concepts in discipline topics widely agreed to be of importance, before and after student learning activities. This approach is well established in examining, for example, student reasoning about basic electricity concepts (e.g. Engelhardt and Beichner, 2004), and a recent publication (Bristow et al., 2012) proposes the use of a multiple-choice diagnostic test to assess student understanding of control engineering concepts.

6. Acknowledgements

The author gratefully acknowledges the comprehensive input of the two anonymous peer reviewers, whose suggestions and advice have encouraged the author in his preparation of the final manuscript.
7. References


