

Briefing Paper on Multiple Choice Assessment

Faculty of Health, Leeds Metropolitan University

Purpose

To advise group heads and other senior staff on the quality assurance aspects of multiple choice for summative and practice assessment.

Summary of Advice

1. MCQ questions are not suitable for all assessments but in contrary to popular belief they can address higher cognitive skills and not just factual recall.
2. Valid questions are hard to write - study time and/or training are required before starting.
3. Bad practice in the creation of questions is hard to un-learn.
4. Creation of questions should include team work and editorial supervision.
5. If large numbers of questions are created from scratch they should pass through the editorial process in small batches.
6. Questions should reach the editor many weeks ahead of the date for delivery.
7. Each faculty group should discuss which question formats and marking schemes are suitable and agree on a short list.
8. Summative assessments should always be invigilated - regardless of the role of computers in the process.
9. In summative assessments responses should usually be collected on sheets designed for automated marking through an optical mark reader.
10. Student responses to questions should always be analysed statistically to validate the questions.
11. Statistical analysis of student responses can be computer automated but interpretation requires skill and training.
12. Even the most experienced writer of questions should expect a proportion of their items to be rejected following statistical analysis.
13. Statistical analysis of responses to MCQ papers can be a powerful tool for targeting the right subject matter for individual and class coaching. MCQ papers are excellent at pinpointing specific misconceptions of the class.

Detailed Advice

1. MCQ questions are not suitable for all assessments but in contrary to popular belief they can address higher cognitive skills and not just factual recall.

Some low level cognitive skill cannot be tested by MCQ - e.g. unprompted recall of facts but some high level skills such as evaluation can, e.g. assertion + reason format questions. Modelling can also be treated - for example, a scenario is described and in each question different parameters to the scenario are given. The student is required to identify the outcome by selecting from multiple options.

A number of examples of such questions are available in the notes on MCQs found on the LTU web site.

2. Valid questions are hard to write - study time and/or training are required before starting.

Although a multiple choice question is very short, its creation is a very demanding process. It requires pinpoint identification of one very specific and measurable learning objective per question and careful avoidance of many possible pitfalls. The skills required are very different from those needed for the creation of teaching materials or the creation of other assessment items. For example, it is quite hard to avoid rewarding quiz-wise students who are otherwise ignorant of the subject with high marks - something that defeats the whole purpose of the test.

The MCQ notes on the LTU web site provide advice on the creation of questions and list many possible pitfalls. Each pitfall has advice on editing the question to eliminate the fault.

3. Bad practice in the creation of questions is hard to un-learn.

If a teacher has gone many years without analysing class performance on MCQ examinations they may be blissfully unaware of the invalidity of the questions they produce. A very common mistake is to think that all the questions in a test are acceptable so long as the mean mark on the test is not too low or too high. In fact this may lead to year on year degradation in the quality of the questions.

A quiz may result in a high mean mark because there are many unintentional clues as to the correct answers. If the question setter is unaware of this he/she may make the intellectual content more difficult in the following year to compensate. This may lower the mean mark but fair ranking of students is not helped - in fact it will be the less quiz-wise and more diligent students whose marks drop the most.

The reverse may also occur - if the wording of the questions is overly complex, ambiguous and confusing then a low mean mark may result. Ranking of students will be based more on ability to decode the language than knowledge of the course content. Without an analysis of the questions the setter may 'dumb down' the questions in an attempt to raise the mean mark. The result is that ranking will depend even less on knowledge of the subject.

4. Creation of questions should include team work and editorial supervision.

A test that has a high proportion of confusing or ambiguous questions can have a very serious demoralising effect on a class of students. A test which clearly rewards students who are ignorant of the subject matter but good at guessing strategies and good at unpicking confusing English can cause a sense of injustice. Teamwork and editorial input will considerably reduce the risks.

5. If large banks of questions are created from scratch they should pass through the editorial process in small batches.

It is quite possible for a question setter to repeat the same fundamentally flawed method in every question they create. Fixing the problem could involve a complete rewrite - this will be detected earlier and more painlessly corrected if the editorial process starts early. A novice question setter (one who lacks experience of having questions validated) would do well to discuss progress with an editor after writing as little as five or six questions.

6. Questions should reach the editor many weeks ahead of the date for delivery.

The process of reviewing and editing MCQ questions is a painstaking task which requires expertise in question setting and expertise in the subject domain of the test. It is time consuming and requires sustained concentration. If a tight deadline is added then the quality of the editorial process will decline.

7. Each faculty group should discuss which question formats and marking schemes are suitable and agree on a shortlist.

Many permutations of question format and marking scheme are possible. Choice of format and marking scheme can have a profound effect on the ranking of students on the assessment and therefore the validity of the examination. Sadly, the decision making process often focuses on the class mean mark and the spread of marks across degree classification bands - i.e. how 'hard' or 'easy' the test is. Validity, or in other words the 'fairness' of the test is far more important and is about proper ranking of students in the class - in a fair test a more knowledgeable and intelligent student scores higher than a less knowledgeable and intelligent student.

The validity of an MCQ assessment depends primarily on the quality of the individual questions. However, it is also affected by the opportunities for quiz-wise students to apply guesswork strategies. The more that quiz-wise students perceive opportunities for applying guesswork the less valid the test is. Validity is also negatively impacted if students who lack guessing strategies are afraid to attempt questions.

Choice of marking scheme should be discussed at the same time as discussing the grading scheme for the whole course because the choice of MCQ marking scheme can dramatically affect the marks awarded to the least able students. For example, if you choose to use multiple true/false questions and set a pass mark of 40% then it will be practically impossible to fail the course no matter how difficult the questions are.

This paper doesn't recommend specific question formats or marking schemes but it is suggested that some degree of standardisation should apply to summative examinations and their corresponding practice examinations. It is recommended that group heads and other interested parties discuss the options with Jon Maber of LTU before making a decision.

8. Summative assessments should always be invigilated - regardless of the role of computers in the process.

It is impossible to detect successful plagiarism on MCQ examinations because you expect able students to submit identical answers. This is obvious to most students. When a student perceives that cheating is easy this leads to the conclusion that most of their fellow students will choose to cheat and therefore they will drop significantly in the class ranking unless they also cheat. Although unproven, one can reasonably assume that some students will become habituated to cheating by this process and continue the practice into other assessment items such as dissertations.

9. In summative assessments responses should usually be collected on sheets designed for automated marking through an optical mark reader.

Optical mark reading (OMR) is preferable to the collection of student response from a computer for a number of reasons;

- It results in a record of student responses which can be referred to in the event of a dispute. The record is created by the student and no alleged software or hardware faults can affect it. Students can and do blame software bugs for low scores on multiple choice exams.
- Normal examination rooms can be used making it easier to timetable them.
- The smooth running of the examination does not depend on the continuous and reliable operation of computer hardware and software for the entire duration of the exam.

There is however a barrier to the adoption of OMR for 100% of summative assessments. Only certain question types are supported by centrally provided services. In order to support a particular question type and its marking scheme there need to be three elements of compatibility; the OMR machine, the software that processes data from the OMR and the printed response sheets. The Leeds Met. OMR hardware and software is compatible with a wide range of multiple choice types and marking schemes but the mass produced stock of response sheets only allows pure MCQ questions with a maximum of five options. LTU is making progress on the creation of software which will print response sheets on a laser printer customised to each examination.

If an examination consists of a combination of MCQ and other question types, for example short written answers, it will be best to provide students with OMR sheets for the MCQ component and normal answer papers for the other component.

If an examination requires that students view multimedia materials it may still be preferable to collect MCQ responses on OMR sheets. This means that a physical record of the responses is produced and it also means that if a computer fails, the student can be immediately moved to another computer with no loss of work.

If examination responses must be collected using a computer it is necessary to operate a 'lock down' browser on the computers to prevent students accessing reference material on the network and communicating with each other.

10. Student responses to questions should always be analysed statistically to validate the questions.

Even the best question setter will produce some poor questions but without a proper item analysis they will never be detected. If the number of invalid questions is high this may impact the class average and the setter may be prompted to create easier (but still invalid) questions thus compounding the underlying problem. Over years the question setter may gradually produce poorer and poorer questions.

Item analysis gives the question setter feedback on which questions are the most and the least effective and helps him/her hone his/her skills. Notes on the subject of item analysis are included in the MCQ documentation on the LTU web site and it is recommended that teachers analyse their MCQ questions with the support of Jon Maber if they are unfamiliar with the process.

11. Statistical analysis of student responses can be computer automated but interpretation requires skill and training.

The mathematics used to analyse students responses to items in an MCQ paper is really quite simple. It's computation by a computer is nearly instantaneous. However, experience in staff development workshop shows that teachers find the process of interpreting the statistical values challenging. This analysis can't be handed over to a statistician because it requires subject domain expertise and direct experience of the teaching that the students have received - i.e. familiarity with the content of the recommended text book and intimate knowledge of the syllabus. There is therefore little option but for course module leaders to spend time learning how to interpret the statistical analysis.

Training is available from LTU in addition to the online documentation.

12. Even the most experienced writer of questions should expect a proportion of their items to be rejected following statistical analysis.

It is usual for new questions to be created each time that a summative assessment is created and it is no shame to the question setter to find that a small proportion fail to stand up to a test for validity after the examination has been marked. It is important that the analysis takes place so that the question setters can continually improve their skills and so that if questions are reused in later years only those that passed the validity test are selected.

One way to improve success rates is to make use of formulaic questions. For example, questions which present a medical condition and then offer a list of optional symptoms. If questions of this type are tried and showed to be valid then the format can be reused with different medical conditions and symptoms. The validity of the new questions is not guaranteed but there is an improved likelihood of success.

13. Statistical analysis of responses to MCQ papers can be a powerful tool for targeting the right subject matter for individual and class coaching.

A huge advantage of item analysis is in the targeting of coaching to the whole class. MCQs are good at identifying gaps in the knowledge of the class but they are excellent at identifying misconceptions - i.e. things which the class erroneously think are correct. An analysis allows the teacher to identify problem areas or even a specific concept in the subject matter which the class needs guidance on. It can also inform the process of course review - identifying areas that require a fresh approach with the next cohort of students.

Rapid feedback on class performance boosts morale and will improve student approval of the course and its teachers; but only if the feedback is meaningful.

When an individual unexpectedly performs poorly on an MCQ test and seeks advice it can be helpful to refer back to individual questions and the responses they gave. If this is done after the test has been through item analysis then the teacher may be better equipped to provide effective coaching.

Other Uses of MCQ

This paper is primarily about the use of MCQ for examinations that count towards a student's grade, **summative** tests and **practice** examinations that mirror them. MCQ test are also useful in other scenarios, for example **formative** tests and **diagnostic** tests. Some of the same issues will apply but others vary - for example, for formative tests it is preferable that students enter responses via a computer rather than OMR because there is no need for a definitive record of the responses and the student will benefit from automated feedback on-screen.

Resources

- The Learning Technology Unit maintains a set of teacher's notes on creation, delivery and analysis of MCQs. This is found in the Resources section of the LTU Intranet site. (The site will soon be reorganised and the document will be found in a new section on computer aided assessment.)
- LTU will run workshops on MCQ assessment at the request of group heads provided that large enough numbers of staff will attend.
- Jon Maber of LTU takes particular responsibility for computer assisted assessment and can provide editorial assistance and coaching to question setters via Email or face to face.
- Jon Maber can also aid teaching staff who want to perform item analysis on their MCQ papers and are unfamiliar with the process.
- The X-stream team can provide technical support for setting up MCQ tests on-line, e.g. for practice assessments.
- The library's staff resource room contains an OMR machine and computer for marking of summative exams performed on optical mark sheets.
- Blank sheets for MCQ examinations are available from the faculty office on level 8 of building D.
- Contact Jon Maber well in advance of an examination if the standard blank sheets are unsuitable and you would like sheets customised to your MCQ questions. These custom sheets can be produced with the student name and number coded on them so the students don't need to do that themselves during the exam.