

Set a multiple choice question (MCQ) examination

David Lowe

In medicine there are few words that are read so critically and on which so much is based

In undergraduate and postgraduate examinations there is little to equal the scrutiny that is given to a multiple choice question. This form of examination has become widespread and generally accepted over the past 20 years,¹ and there can be few of us that have not sat, composed, or organised one. For examiners there are several attractive features: objectivity, ease of scoring when computer marking facilities are available, and simplicity of interpreting the results.² For the undergraduate and postgraduate students who have to answer such examinations things are not so easy. There are different terrors from writing an essay or taking a viva examination, and candidates always protest that some of the questions are ambiguous or incomprehensible. They are sometimes right.

Organising a multiple choice examination

Organising a multiple choice examination is fairly straightforward.³ You will need to think about four main points: the standard and scope of the examination, the number of questions you wish to ask, the time you have available for the examination, and the source of your questions. This last point may be the most difficult. It is usually easy to decide the total number of questions and the proportion in any subsections of the examination. The acceptable number of questions per hour will depend on the standard of the examination and the complexity of the questions. For undergraduate examinations a reasonable rate is one minute per question and an additional 10 minutes for every 50 questions, but some postgraduate examining bodies set only 30 questions per hour.

Styles of multiple choice question

There are many styles that you can use to set multiple choice examinations. The simplest is a list of questions that require a yes or no answer, though this is not strictly multiple choice. There is the "pick one of the following" sort (also called "one from five"), which has a question and five possible answers, of which four are wrong. Other styles that have been used include pairing from two lists ("match each of the diseases 1-5 with the treatments A-E"); arranging a set of five choices into a logical sequence; and problem solving of clinical history and investigations for interpretation. There is a complicated relation analysis style in which the candidate has to decide whether an assertion and a reason ("mesothelioma is a characteristic tumour of asbestos workers BECAUSE asbestos penetrates mesothelial cells") are true and conse-

Box 1—Example of acceptable question

IgG (immunoglobulin G):

- A Fixes complement
- B Crosses the placenta
- C Is formed as part of a secondary immune response
- D Triggers leukotriene production
- E Takes part in the allergic response

Box 2—Example of acceptable question

B lymphocytes are:

- A Phagocytic
- B Precursors of plasma cells
- C Able to carry immunoglobulin on their surfaces
- D Found in lymphoid follicles
- E Stimulated into mitosis by phytohaemagglutinin

quential, true but not a consequence, or false, but happily this style is seldom used.

The basic format in most multiple choice examinations is the "determinate response" or "multiple true-false" style. There is an introductory statement or *stem* followed by a choice of five *options* or *completions*. The stem and an option together are called an *item*, and any of these may be true or false. (Options have been called *responses* in some publications, but as the stem should be an incomplete statement and not a question its predicate should not be called a response. It is the candidate who makes a response to an item.) False items are called *distractors*. The candidate has to decide whether each item is true or false and in medicine is usually allowed to respond with don't know, on the grounds that recognition of uncertainty is important for a doctor. The decision is marked as correct, incorrect, or zero, usually by computer.

Writing multiple choice questions

Writing multiple choice questions is a laboriously acquired art, and it is much easier to advise how not to write them than to give a simple way of writing them. A few guidelines are worth considering. The questions should be

- Relevant
- Short
- Understandable
- Discriminant.

That is, they should be worth asking and based on the right subject matter, be concise and clear, and differentiate acceptable from unacceptable candidates. Boxes 1 and 2 give examples of acceptable questions and boxes 3 and 4 unacceptable ones. Data for the questions may be derived from any source; textbooks, review articles, and case reports that have a literature review are often very useful. The latest medical research is not usually suitable for multiple choice questions unless the data are widely publicised.

To compose a multiple choice question you could make a list of the points from an article that you think are essential facts or core knowledge and then think of a stem that leads the question into the right area. Distractors are usually more difficult to invent than true items, and the mark of a good question is often the quality of the distractors. You could try substituting a similar anatomical site (small for large bowel), an attribute of a related substance or condition (IgM: A, crosses the placenta), or a recognised misconception (Testicular teratoma: A, typically metastasises to inguinal lymph nodes), but you must ensure that the distractor items really are false, and this can be very difficult.

Department of
Histopathology, St
Bartholomew's Hospital
Medical College, London
EC1A 7BE
David Lowe, MD, senior
lecturer

BMJ 1991;302:780-2

Relevance

The questions should test important knowledge. Many medical facts fit well into a multiple choice format but may be quite unimportant in the theory or practice of a specialty. For example, in a clinical examination a question can easily be set asking which blood group is commonest in patients with gastric carcinoma, but knowledge of this is of very limited diagnostic or therapeutic use. Similarly, do general surgeons need to know whether corticotrophin releasing hormone is or is not a tripeptide?

The balance of topics should be checked. When an examination is in several parts and the questions are submitted by different parties it is easy for a topic to be duplicated. For example, the histopathology and microbiology parts of a pathology multiple choice examination could ask very similar questions about tuberculosis. When the questions are taken from a bank you should check the range of topics that they cover. You may be provided with 45 questions on neoplasia and five on inflammatory and metabolic conditions rather than an even spread. If the scope of the subject matter is not controlled the candidates' marks may be impossible to interpret.

Brevity and clarity

The stem and the five options connected to it should be unambiguous, readily understood, and as short as possible. The stem can be a short statement or outline a clinical picture: "Nasal polyps are:" or "A patient who has recently returned from Africa presents with rigors. Urgent investigations should include:" Stems that are only one word usually need long options. This is not necessarily a problem (boxes 1 and 2), but there is a danger that the determinate response style of a multiple choice question can be weaseled into a set of five straight true/false questions only faintly related to each other: "Pyrexia: A, is diagnostic of appendicitis; B, results from the effect of enzymes on the limbic system; C, of unknown origin is investigated with blood cultures." This style of question takes longer to read and answer, and if there are more than a few the time available for the examination may have to be extended.

There should be no double negatives, and single negatives should be avoided if possible (box 3, D). Try to include in the stem as many parts of the options as you can—for example, if the stem is "Arthritis:" and all of the options start with "is a..." you could rewrite the stem as "Arthritis is a:" (see box 2, stem). Woolly stems such as "Concerning the obturator foramen:" are old fashioned and can usually be turned into crisper phrases.

Some words, such as always, never, and only are contentious in an exact science like medicine and should not be used.² Others, such as often, commonly, rarely, sometimes, frequently, and may should be used very carefully—how frequent is a condition that is common or rare (box 3, A)? Items that have the verb "may" have a high chance of being correct (box 4, C and D). Exact percentages for incidence, prognosis,

Box 3—Example of unacceptable question

Cystic fibrosis:

- A Is rarely a disorder of endocrine glands
- B Is inherited as an autosomal dominant condition with an abnormality of chromosome 5
- C Patients have a reduced sodium chloride in sweat
- D Is not associated with intestinal obstruction
- E Results in biliary cirrhosis in 10% of cases

Box 4—Example of unacceptable question

Endocarditis:

- A α Haemolytic strains of enterococci are the commonest bacterial cause
- B Requires three sets of blood cultures to isolate the pathogens in most cases of culture positive endocarditis
- C Caused by streptococci may be treated with a combination of penicillin and gentamicin
- D Associated recurrent fever may frequently be due to superinfection of a valve by antibiotic resistant bacteria
- E Non-bacterial causes of endocarditis are very rare

etc, should be avoided (box 3, E); they are too dependent on the latest reference and may vary widely from source to source. You could argue for using a percentage if the item is obviously wrong—"Atheroma: A, occurs in 1% of the adult population in GB"—but if it is obvious it is unlikely to be a very discriminatory question. Similarly, absolute values for biochemical and peak flow measurements, etc, should be avoided. It is better to say "...of three times the normal maximum concentration" so that differences in reference ranges can be taken into account.

The stem and option should read logically. It is easy to write items that look adequate but are not proper English (box 4, A and E) or do not make sense (box 3, B)—for example, "Pulmonary microemboli: A, impairs gas exchange" or "Sickle cell anaemia: B, has two bands on electrophoresis." Tenses should be the same throughout; the present tense is the most usual. There should be no mutually exclusive or inclusive options, such as "Serum α fetoprotein concentration is raised in: A, all malignant tumours; B, adenocarcinoma of pancreas." If a word in a stem is used as a noun for most of the options try not to let it become adjectival: "Bilirubin: A, is a breakdown product of myoglobin; B, concentration in the serum increases in cirrhosis." (Box 3, C). Useful standard terms include "Characteristic features of X are seen in," "recognised features of X include," and "X is typical of." Characteristic can be defined as a textbook feature without which the diagnosis may be in doubt and recognised and typical as referring to established facts, but all three adjectives have essentially similar meanings.

The stem should not ask a question. The answer to "Which of the following..." is logically this one or that one rather than true or false. Another problem with this construction is that the verb can indicate that one ("...of the following is...") or more than one ("...of the following are...") of the options is correct.

Stems with a double statement can be ambiguous. For example, "Causes of hyponatraemia and hyperkalaemia include:"—does the stem mean hyponatraemia and hyperkalaemia occurring together, or is it referring to causes of hyponatraemia and of hyperkalaemia, separately? Options with two parts similarly cause problems: "Gastric carcinoma: A, characteristically occurs in patients with pernicious anaemia because of delay in gastric emptying" effectively asks two questions (box 3, B; box 4).

Having laboriously constructed your questions and found them to be faultlessly logical, crystal clear, and in impeccable English, you then come to the hard part. It is essential that new questions are criticised by as many people as possible. These should include fellow examiners and, for undergraduate examinations, examiners in other disciplines. Previous candidates in an examination can also provide useful feedback. If there is more than one way of interpreting a point this should become apparent, and there may be disagreement about the value of asking some of the questions.

If your questions survive scrutiny by a panel of critical examiners it is likely that they will be good questions.

Discrimination and assessment of results

All multiple choice questions should be reviewed by the examiners before each use to check that the questions are acceptable and the answers are correct. Ideas in medicine change with new discoveries, and out of date questions can be very confusing. Questions that consistently fail to discriminate between strong and weak candidates should be revised or excluded. This analysis can also indicate whether a bank of questions has been broken because the questions have become public property, a problem especially of small banks. Simple changes to stems may be all that are needed to revive poor questions or to reconstruct the bank.

When a multiple choice examination is being used for the first time as part of an assessment the results can be compared with previous candidates' performances in essay questions and vivas. A multiple choice examination may be run in parallel with essays and vivas so that there is an internal control. New questions could be added to the end of a series of established questions to see how they compare.

It is usual for at least one item in a five part question to be true: in practice there are remarkably few questions in which all of the items are false. The first item is more likely to be true than false, and there are likely to be more true items than false ones. These findings can affect the statistical assessment of the results of the examination. When you compose multiple choice questions you should think of including one or two in which the items are all true or all false and ensure that the options of the others are randomised.⁵

The simplest marking system is to give one mark for each correct answer. This was done in the early days of multiple choice examinations but encouraged guessing as there was no penalty for an incorrect answer. To overcome this, negative marking was introduced, in which an incorrect answer is penalised by deduction of a mark. A "don't know" answer does not lose the candidate a mark.

For each of the five items of a multiple choice question the computer marking the candidates' papers calculates a *facility index* and a *discrimination index*. These values are not absolute and apply only to the use of the question in a particular examination. When used again in another examination the values of the indices may change, though on the whole there is good correlation between the responses of different groups of candidates. The facility index is simply the proportion of candidates answering the item correctly—if two thirds give the right answer the facility index is 0.67. Items with a high facility index ("easy" questions) can be acceptable when there are others with lower indices as part of the set of five. The discrimination index measures how well an item distinguishes between strong and weak candidates. From the examination as a whole a percentage of the good and bad candidates is taken: some examining bodies take the top 27% and the bottom 27%, others take the top and bottom halves. Then for each item the number of the correct answers given by the bottom group is subtracted from the number given by the top and the result is divided by the percentage of students in the groups (27, 50, or whatever percentage was used). In most cases the result will be a positive number. If the discrimination index for an item is low the item should be checked; if the index is negative (the poor students did better than the best students) then the item is usually poorly constructed and the answer provided may be wrong.

Public relations

A candidate's performance in a multiple choice examination should reflect his or her knowledge of medicine and not the ability to understand linguistic nuance. If use of English is to be tested essay questions or vivas are a better way to do this. Sample multiple choice questions and brief instructions on the way to deal with them should be available for candidates who are unaccustomed to this form of examination, so that results can be compared, for example, when the examination is held in different countries.

If the equipment that will be used to assess the computer sheets reacts only to pencil marks and not to ink the candidates should be warned of this. Candidates may choose to write their answers in the first instance on the question paper and then transfer them to the computer marking sheet; they should be told clearly that this must be done within the allotted time of the examination and that extra time to transfer answers will not be allowed. Even with this warning, and even when it is repeated near the end of the examination, it is surprising how often candidates fail to enter their answers on the computer sheet in the available time.

Feedback from the candidates about the questions can be very informative. Because of the extra time needed feedback is unlikely to be possible in qualifying examinations, when conditions must be strictly controlled, but in end of term examinations a blank sheet may be provided and an extra 10 minutes allowed for candidates to write their comments. It is unlikely that this will stop candidates complaining, but at least they will know you care.

In conclusion the multiple choice question paper has become established as a part of undergraduate and postgraduate medical examinations, and whether we like them or not we are obliged, as Anderson suggested, to "write good questions, evaluate them and appreciate their invaluable, if limited, place in assessment."³ The first takes practice, the second is relatively easy, and the third may be difficult but should not be impossible.

- 1 Anderson J. *The multiple choice question in medicine*. London: Pitman Medical, 1976.
- 2 Slade PD, Dewey ME. Role of grammatical clues in multiple choice questions: an empirical study. *Medical Teacher* 1983;5:146-8.
- 3 Lennox B. *Hints on the setting and evaluation of multiple choice questions of the one-from-five type*. Dundee: Association for the Study of Medical Education, 1974.
- 4 Anderson J. The MCQ controversy—a review. *Medical Teacher* 1981;3:150-6.
- 5 Anderson J. For multiple choice questions. *Medical Teacher* 1979;1:37-42.

Update box for *Oxford Handbook of Clinical Medicine* (2nd ed), p 240

Avoiding melarsoprol induced encephalopathy

Lethal encephalopathy occurs in up to 10% of patients with trypanosomiasis undergoing treatment with melarsoprol. It presents with abnormal behaviour, seizures, and coma and is partly preventable by giving oral prednisolone before the first injection of melarsoprol.¹ The dose of prednisolone is 1 mg/kg/24 h, up to 40 mg. This should be continued throughout the first two series of melarsoprol injections. Whenever possible, discuss the management of the patient with an experienced colleague. — J M LONGMORE

Principal source

- 1 Pepin J, Milord F, Guern C, Mpia B, Ethier L, Mansina D. Trial of prednisolone for prevention of melarsoprol-induced encephalopathy in Gambiense sleeping sickness. *Lancet* 1989;i:1246-50.

Anyone may submit an update box; all boxes are peer reviewed.