GUIDELINES FOR THE CONSTRUCTION OF MULTIPLE CHOICE QUESTIONS TESTS

Mohammed O. Al-Rukban, *DPHC, ABFM, SBFM*
Department of Family & Community Medicine, College of Medicine and King Khalid University Hospital, King Saud University, Riyadh, Saudi Arabia

تعتبر أسئلة الاختيار المتعدد من أكثر وسائل التقويم استخداماً في التعليم الطبي. فهي خيار جيد لقياس معرفة المتدربين ومستوى فهمهم وقدراتهم العملية على حل المشكلات والحالات الطبية. وتهدف هذه الورقة إلى إعانة المهتمين بتقويم الطلاب والمتدربين وكذلك المنشغلين بالتعليم الطبي على معرفة الخطوات العلمية والصحيحة لتكوين أسئلة الاختيار المتعدد من النوع (Single Best Option) والنوع (Extended Matching Item)، مع بعض القوالب التي يمكن استخدامها في تكوين هذه الأسئلة.

الكلمات المرجعية: التقويم، أسئلة الاختيار المتعدد ، الطلاب، التدريب.

Multiple Choice Questions (MCQs) are generally recognized as the most widely applicable and useful type of objective test items. They could be used to measure the most important educational outcomes – knowledge, understanding, judgment and problem solving. The objective of this paper is to give guidelines for the construction of MCQs tests. This includes the construction of both "single best option" type, and "extended matching item" type. Some templates for use in the "single best option" type of questions are recommended.

Key Words: Multiple Choice Questions (MCQs), Assessment, Extended matching Item, Evaluation, Students, and Training.

INTRODUCTION

In recent years, there has been much discussion about what should be taught to medical students and how they should be assessed. In addition, highly publicized instances of the poor performance of medical doctors have fuelled the drive to find a way for ensuring that qualified doctors achieve and maintain appropriate knowledge, skills and attitudes throughout their working lives.[1]

Selecting an assessment method for measuring students' performance remains a daunting task for many medical institutions.[2] Assessment should be educational and formative if it is going to promote appropriate learning. It is important that individuals learn from any assessment process and receive feedback on which to build their knowledge and skills. It is also important for an assessment to have a summative function to demonstrate competence.[1]

Assessment may act as a trigger, informing examinees what instructors really regard as important[3] and the value they attach to different forms of knowledge and ways of thinking. In fact, assessment has been identified as possibly the single most potent influence on student learning; narrowing students' focus only on topics to be tested on (i.e. what is to be studied) and shaping their learning approaches (i.e. how it is going to be studied).[4] Students have been found to differ in the quality of their learning when instructed to focus either on factual details or on the assessment of evidence.[5] Furthermore, research has reported that changes in assessment methods have been found to influence medical students to alter their study activities.[4] As methods of assessment drive learning in medicine and other disciplines,[1] it is important that the assessment tools test the attributes required of students or professionals undergoing revalidation. Staff subsequently, redesign their methods of assessment to ensure a match between assessment forms and their educational goals.[6]

Correspondence to:
Dr. Mohammed O. Al-Rukban, P.O. Box 91678, Riyadh 11643, Saudi Arabia - E-mail: mrukban@ksu.edu.sa

Methods of assessment of medical students and practicing doctors have changed considerably during the last 5 decades.[7] No single method is appropriate, however, for assessing all the skills, knowledge and attitudes needed in medicine, so a combination of assessment techniques will always be required.[8-10]

When designing assessments of medical competencies, a number of issues need to be addressed; reliability, which refers to the reproducibility or consistency of a test score, validity, which refers to the extent to which a test measures what it purports to measure,[11,12] and standard setting which defines the endpoint of the assessment.[1] Sources of the evidence of validity are related to the content, response process, internal structure, relationship to other variables, and consequences of the assessment scores.[13]

Validity requires the selection of appropriate test formats for the competencies to be tested. This invariably requires a composite examination. Reliability, however, requires an adequate sample of the necessary knowledge, skills, and attitudes to be tested.

However, measuring students' performances is not the sole determinant for choosing an assessment method. Other factors such as cost, suitability, and safety have profound influences on the selection of an assessment method and, most probably, constitute the major reason for inter-institutional variations for the selection of assessment methods as well success rates.[14]

Examiners need to use a variety of test formats when organizing test papers; each format being selected on account of its strength as regards to validity, reliability, objectivity and feasibility.[15]

For as long as there is a need to test knowledge in the assessment of doctors and medical undergraduates, multiple choice questions (MCQs) will always play a role as a component in the assessment of clinical competence.[16]

Multiple choice questions were introduced into medical examinations in the 1950s and have been shown to be more reliable in testing knowledge than the traditional essay questions. It represents one of the most important well-established examination tools widely used in assessment at the undergraduate and postgraduate levels of medical examinations. The MCQ is an objective question for which there is prior agreement on what constitutes the correct answer. This widespread use may have led examiners to use the term MCQ as synonym to an objective question.[15] Since their introduction, there have been many modifications to MCQs resulting in formats.[16] Like other methods of assessment, they have their strengths and weaknesses. Scoring of the questions is easy and reliable, and their use permits a wide sampling of student's knowledge in an examination of reasonable duration.[15-19] MCQ-based exams are also reliable because they are time-efficient and a short exam still allows a breadth of sampling of any topic.[19] Well-constructed MCQs can also assess taxonomically higher-order cognitive processing such as interpretation, synthesis and application of knowledge rather than the test of recall of isolated facts.[20] They could test a number of skills in addition to the recall of factual knowledge, and are reliable, discriminatory, reproducible and cost-effective. It is generally, agreed that MCQs should not be used as a sole assessment method in summative examinations, but alongside other test forms. They are designed to broaden the range of skills to be tested during all phases of medical education, whether undergraduate, postgraduate or continuing.[21]

Though writing the questions requires considerable effort, their high objectivity makes it possible for the results to be released immediately after marking by anyone including a machine.[15,18] This facilitates the computerized analysis of the raw data and allows the examining body to compare the performance of either the group or an individual with that of past candidates by the use of discriminator questions.[22] Ease of marking by computer makes MCQs an ideal method for assessing the knowledge of a large number of candidates.[16, 22]

However, a notable concern of many health professionals is that they are frequently faced with the task of constructing tests with little or no experience or training on how to perform this task. Examiners need to spend considerable time and effort to produce satisfactory questions.[15]

The objective of this paper is to describe guidelines for the construction of two common MCQs types: the "single best option" type, and "extended matching item" type. Available templates for the "single best option" type will be discussed.

Single Best Option
The first step for writing any exam is to have a blueprint (table of specifications). Blueprinting is the planning of the test against the learning objectives of a course or competencies essential to

a specialty.[1] A test blueprint is a guide used for creating a balanced examination and consists of a list of the competencies and topics (with specified weight for each) that should be tested on an examination, as in the example presented in Table 1.

If there is no blueprint, the examination committee should decide on the system to be tested by brainstorming to produce a list of possible topics/themes for question items. For example, abdominal pain, back pain, chest pain, dizziness, fatigue, fever, etc,[23] and then select one theme (topic) from the list. When choosing a topic for a question, the focus should be on one important concept, typically a common or a serious and treatable clinical problem from the specialty. After choosing the topic, an appropriate context for the question is chosen. The context defines the clinical situation that will test the topic. This is important because it determines the type of information that should be included in the stem and the response options. Consider the following example: (Topic= Hypertension; Context= Therapy).

The basic MCQ model comprises a stem and a lead-in question followed by a number of answers (options).[19] The option which matches the key in a MCQ is best called "the correct answer"[15] and the other options are the "distracters".

Table 1: *Example of a table of specifications (Blueprint) based on the context, for Internal Medicine examination*

Branches	HM	Mech-anism	Dx	Manage-ment	Total
Cardiovascular	3	4	9	9	25
Nephrology	1	3	4	4	12
Respiratory	2	3	8	9	22
GIT	1	3	5	4	13
Rheumatology	3	2	3	5	13
Dermatology	1	0	3	2	6
Hematology	1	3	3	4	11
Infectious	1	2	9	5	17
Neurology	1	5	8	7	21
Total	**14**	**25**	**52**	**49**	**140**

HM=Health maintenance

For writing a single best option type of MCQs, as shown in Appendix 1, it is recommended that the options are written first.[23] A list of possible homogeneous options based on the selected topic and context is then generated. The options should be readily understood and as short as possible.[18] It is best to start with a list of more than five options (although only five options are usually used in the final version). This allows a couple of 'spares', which often come in handy!

It is important that this list be HOMOGENOUS (i.e. all about diagnoses, or therapeutics, lab investigations, complications… etc)[23] and one of the options selected as the key answer to the question.

A good distracter should be inferior to the correct answer but should also be plausible to a non-competent candidate.[24] All options should be true and contain facts that are acceptable to varying degrees. The examiner would ask for the most appropriate, most common, least harmful or any other feature which is at the uppermost or lowermost point in a range. It needs to be expressed clearly that only one answer is correct. A candidate's response is considered correct if his/her selection matches the examiner's key.[15]

When creating a distracter, it helps to predict how an inexperienced examinee might react to the clinical case described in the stem.[24]

A question stem is then written with lead-in statement based on the selected correct option. Well-constructed MCQs should test the application of medical knowledge (context-rich) rather than just the recall of information (context-free). Schuwirth et al,[25] found that context-rich questions lead to thinking processes which represent problem solving ability better than those elicited by context-free questions. The focus should be on problems that would be encountered in clinical practice rather than an assessment of the candidate's knowledge of trivial facts or obscure problems that are seldom encountered. The types of problems that commonly encountered in one's own practice can provide good examples for the development of questions. To make testing both fair and consequentially valid, MCQs should be used strategically to test important content, and clinical competence.[19]

The clinical case should begin with the presentation of a problem and followed by relevant signs, symptoms, results of diagnostic studies, initial treatment, subsequent findings, etc. In essence, all the information that is necessary for a competent candidate to answer the question should be provided in the stem. For example:

Age, sex (e.g., a 45-year-old man).
Site of care (e.g. comes to the emergency department).
Presenting complaint (e.g. because of a headache).
Duration (e.g. that has continued for 2 days).
Patient history (with family history).
Physical findings.
+/- Results of diagnostic studies.

+/- Initial treatment, subsequent findings, etc.

The lead-in question should give clear directions as to what the candidate should do to answer the question. Ambiguity and the use of imprecise terms should be avoided.[16,18] There is no place for trick questions in MCQ examinations. Negative stems should be avoided, as should double negatives. Always, never and only are obviously contentious in an inexact science like medicine and should not be used.[16, 18] Consider the following examples of lead-in questions:

Example 1: Regarding myocardial infarction.

Example 2: What is the most likely diagnosis? Note that for Example 1, no task is presented to the candidate. This type of lead-in statement will often lead to an ambiguous or unfocused question. In the second example, the task is clear and will lead to a more focused question. To ensure that the lead-in question is well constructed, the question should be answerable without looking at the response options. As a check, the response options should be covered and an attempt made to answer the question.

Well constructed MCQs should be written at a level of difficulty appropriate to level of the candidates. A reason often given for using difficult questions is that they help the examiner to identify the `cream' of the students. However, most tests would function with greater test reliability when questions of medium difficulty are used.[26] An exception, however, would be the assessment of achievement in topic areas that all students are expected to master. Questions used here will be correctly answered by nearly all the candidates and consequently, will have high difficulty index values. On the other hand, if a few candidates are to be selected for honours, scholarships, etc., it is preferable to have an examination of the appropriately high level of difficulty specifically for that purpose. It is important to bear in mind that the level of learning is the only factor that should determine the ability of a candidate to answer a question correctly.[15]

The next step is to reduce the list of option to the intended number of options which is usually five options (including, of course, the correct answer).

Lastly, the option list is to be arranged into a logical order to reduce guessing and avoid putting the correct answer in habitual location (e.g. using alphabetical order will make it possible to avoid choosing options B or C as key answers more frequently).

The role of guessing in answering MCQs has been debated extensively and a variety of approaches have been suggested to deal with the candidate who responds to questions without possessing the required level of knowledge.[27-29] A number of issues need closer analysis when dealing with this problem. Increasing the number of questions in a test paper will reduce the probability of passing the test by chance.[15]

Once the MCQs have been written, they should be criticized by as many people as possible and they should be reviewed after their use.[16,18] The most common construction error encountered is the use of imprecise terms. Many MCQs used in medical education contain undefined terms. Furthermore, there is a wide range of opinions among the examiners themselves about the meanings of such vague terms.[30] The stem and options should read logically. It is easy to write items that look adequate but do not constitute proper English or do not make sense.[18]

When constructing a paper from a bank of MCQs, care should be taken to ensure that there is a balanced spread of questions across the subject matter of the discipline being tested.[16] A fair or defensible MCQ exam should be closely aligned with the syllabus; be combined with practical competence testing; sample taken broadly from important content and be free from construction errors.[19]

Extended Matching Items (EMIs)

Several analytic approaches have been used to obtain the optimal number of response options for multiple-choice items.[31-35] Focus has shifted from traditional 3–5 branches to larger numbers of branches. This may be 20–30 in the case of extended-matching questions (EMIs), or up to 500 for open-ended or 'uncued' formats.[36] However, the use of smaller numbers of options (and more items) results in a more efficient use of testing time.[37]

Extended-matching items are multiple choice items organized into sets that use one list of options for all items in the set. There is a theme, an option list, a lead-in statement and at least two item stems. A typical set of EMIs begins with an option list of four to 26 options; more than ten options are usually used. The option list is followed by two or more patient-based items requiring the examinee to indicate a clinical decision for each item. The candidate is asked to match one or more options to each item stem.

Extended matching items have become popular in such specialties as internal and family medicine because they can be used to test diagnostic ability and clinical judgment.[15] Its use likely to increase in postgraduate examinations as well as in undergraduate assessment.[21] Computer-based extended matching items have been used for in-course continuous assessment.[38]

EMIs are more difficult, more reliable, more discriminating, and capable of reducing the testing time. In addition, they are quicker and easier to write than other test formats.[39, 40] Over the past 20 years, multiple studies have found that EMI-based tests are more reproducible (reliable) than other multiple-choice question (MCQ) formats aimed at the assessment of medical decision making.[20,41,42] There is a wealth of evidence that EMIs are the fairest format.[19]

Another more recent development is uncued questions where answers are picked from a list of several hundred choices. These have been advocated for use in assessing clinical judgment,[43] but extended matching questions have surprisingly been shown to be as statistically reliable and valid as uncued queries.[20,41,44]

Extended matching questions overcome the problem of cueing by increasing the number of options and are a compromise between free-response questions and MCQs. This offers an objective assessment that is both reliable and easy to mark.[45-47]

Nevertheless, MCQs have strengths and weaknesses and those responsible for setting MCQ papers may consider investigating the viability and value of including some questions in the extended matching format. Item writers should be encouraged to use the EMI format with a large number of options because of the efficiencies this approach affords in item preparation.[20,39]

For the construction of EMIs the following steps are suggested (Appendix 2):

Step 1: The selection of the system, the context, and the theme should be based on a blueprint. Otherwise, the following sequence should be followed:
- Decide on the system or specialty (e.g. Cardiovascular, Respiratory, Paediatrics, Obstetric and Gynaecology, Gastroenterology, etc).
  Example: Respiratory system.
- Decide on the context (e.g. Diagnosis, laboratory investigations, aetiology, therapy or prognosis).
  Example: Laboratory investigations.
- Choose a theme (e.g. Fatigue, antibiotic, enzymes, diet, etc).
  Example: Respiratory Tract Infection.

Step 2: Write the lead-in statement.
Example: Which is the best specimen to send to the Microbiology laboratory for confirmation of the diagnosis?

Step 3: Prepare the options: Make a list of 10-15 homogenous options. For example, they should all be diagnoses, managements, blood gas values, enzymes, prognoses, etc, and each should be short (normally only one or two words).
Example of options:
− *Sputum bacterial culture.*
− *Nasal swab.*
− *Blood for C-reactive protein.*
− *Blood culture.*
− *Perinasal swab.*
− *Cough swab.*
− *Throat swab.*
− *Bronchioalveolar lavage.*
− *Urine for antigen detection.*
− *Single clotted blood specimen.*

Step 4: Select two to three of the options as the correct answers (keys).
Example:
− *Sputum bacterial culture.*
− *Blood culture.*
− *Bronchioalveolar lavage.*

Step 5: The question stems or scenarios: Write two to three vignettes (case scenarios) that suit the selected options, to form the question stem. The scenarios should not be overly complex and should contain only relevant information. This should normally be between two and five sentences in length. For the questions:
a. Use patient scenarios.
b. Include key patient information.
c. Structure all similar scenarios in one group (do not mix adult and paediatric scenarios).
d. Scenarios should be straightforward.
Scenarios for most of the options for possible use in future examinations can be written.
Example:
- *A 21-year-old man severely ill with lobar pneumonia.*
- *A 68-year-old woman with an exacerbation of COPD.*

- *A 60-year-old man with strongly suspected TB infection on whom three previous sputum specimens have been film (smear) negative.*

N.B: The response options are first written and then the appropriate scenario built for each one.

Step 6: Ensure validity and discrimination:
Look at other options in the provisional list and delete any that suits the written case scenario or that is clearly wrong or plausible.

Step 7: Reduce the option list to the intended number e.g. (10-15 options for a two case scenarios).

Step 8: Review the questions and ensure that there is only one best answer for each question. Ensure that there are at least four reasonable (plausible) distracters for each scenario and ensure that the reasons for matching are clear. It is advisable to ask a colleague to review the EMIs without the answers. If the colleague has difficulty, modify the option list or scenario as appropriate.

The following is a sample:
Respiratory Tract Infection
For the following patients, which is the best specimen to send to the Microbiology laboratory for confirmation of the diagnosis?
A 21-year-old man severely ill with lobar pneumonia. (d)
A 6- year-old woman with an exacerbation of COPD. (a)
A 60-year-old man with strongly suspected TB infection on whom three previous sputum specimens have been film (smear) negative. (h)

Options:
a. Sputum bacterial culture.
b. Nasal swab.
c. Blood for C-reactive protein.
d. Blood culture.
e. Perinasal swab.
f. Cough swab.
g. Throat swab.
h. Bronchioalveolar lavage.
i. Urine for antigen detection.
j. Single clotted blood specimen.

It is advisable for the writer of MCQ to use templates for the construction of a single best option MCQ in both basic sciences and physician (clinical) tasks.[48]

Although the topics in basic sciences could be tested by recall type MCQs, as was discussed earlier, case scenario questions are preferable. Therefore, the focus here will be on this type of questions. The components of patient vignettes for possible inclusion were also described earlier.

Patient Vignettes
- A (patient description) has a (type of injury and location). Which of the following structures is most likely to be affected?
- A (patient description) has (history findings) and is taking (medications). Which of the following medications is the most likely cause of his (one history, physical examination or lab finding)?
- A (patient description) has (abnormal findings). Which [additional] finding would suggest/suggests a diagnosis of (disease 1) rather than (disease 2)?
- A (patient description) has (symptoms and signs). These observations suggest that the disease is a result of the (absence or presence) of which of the following (enzymes, mechanisms)?
- A (patient description) follows a (specific dietary regime); which of the following conditions is most likely to occur?
- A (patient description) has (symptoms, signs, or specific disease) and is being treated with (drug or drug class). The drug acts by inhibiting which of the following (functions, processes)?
- (Time period) after a (event such as trip or meal with certain foods), a (patient or group description) became ill with (symptoms and signs). Which of the following (organisms, agents) is most likely to be found on analysis of (food)?
- Following (procedure), a (patient description) develops (symptoms and signs). Laboratory findings show (findings). Which of the following is the most likely cause?

Sample Lead-ins and Option Lists
- Which of the following is (abnormal)?
  Option sets could include sites of lesions; list of nerves; list of muscles; list of enzymes; list of hormones; types of cells; list of neurotransmitters; list of toxins, molecules, vessels, and spinal segments.
- Which of the following findings is most likely or most important?

Option sets could include a list of laboratory results; list of additional physical signs; autopsy results; results of microscopic examination of fluids, muscle or joint tissue; DNA analysis results, and serum levels.
- Which of the following is the most likely cause?
  Option sets could include a list of underlying mechanisms of the disease; drugs or drug classes that might cause side effects; toxic agents; hemodynamic mechanisms, viruses, and metabolic defects.

Items Related to Physician Tasks[48]

Diagnosis
The classic diagnosis item begins with a patient description (including age, sex, symptoms and signs and their duration, history, physical findings on exam, findings on diagnostic and lab studies) and ends with a question:
- Which of the following is the most likely diagnosis?
- Which of the following is the most appropriate next step in diagnosis?
- Which of the following is most likely to confirm the diagnosis?

Management:
Questions to ask include:
- Which of the following is the most appropriate initial or next step in patient care?
- Which of the following is the most effective management?
- Which of the following is the most appropriate pharmacotherapy?
- Which of the following is the first priority in caring for this patient? (eg, in the emergency department).

Health and Health Maintenance:
The following lead-ins are examples of those used in this category:
- Which of the following immunizations should be administered at this time?
- Which of the following is the most appropriate screening test?
- Which of the following tests would have predicted these findings?
- Which of the following is the most appropriate intervention?
- For which of the following conditions is the patient at greatest risk?

- Which of the following is most likely to have prevented this condition?
- Which of the following is the most appropriate next step in the management to prevent [morbidity/mortality/disability]?
- Which of the following should be recommended to prevent disability from this injury/condition?
- Early treatment with which of the following is most likely to have prevented this patient's condition?
- Supplementation with which of the following is most likely to have prevented this condition?

Mechanisms of Disease:
Begin your mechanism items with a clinical vignette of a patient and his/her symptoms, signs, history, laboratory results, etc., then ask a question such as one of these:
- Which of the following is the most likely explanation for these findings?
- Which of the following is the most likely location of the patient's lesion?
- Which of the following is the most likely pathogen?
- Which of the following findings is most likely to be increased/decreased?
- A biopsy is most likely to show which of the following?

REFERENCES
1. Wass V, Van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. Lancet 2001; 357(9260): 945-9.
2. Walubo A, Burch V, Parmar P, Raidoo D, Cassimjee M, et al. A Model for Selecting Assessment Methods for Evaluating Medical Students in African Medical Schools. Acad Med 2003; 78(9): 899-906
3. Crooks TJ, Mahalski PA. Relationships among assessment practices, study methods, and grades obtained. Research and Development in Higher Education 1985;8:234-40.
4. Scouller KM, Prosser M. Students' Experiences in Studying for Multiple Choice Question Examinations. Studies in Higher Education. Carfax Publishing Company 1994;19(3):267-73.
5. Biggs JB. Individual differences in study processes and the quality of learning outcomes. Higher Education 1979;8: 381-94.

6. Newble DI, Jaeger K. The effect of assessments and examinations on the learning of medical students. Med Educ 1983;17(3):165-71.
7. Schuwirth LW, Van der Vleuten CP. Changing education, changing assessment, changing research? Med Edu 2004; 38(8):805-12.
8. Collins JP, Gamble GD. A multi-format interdisciplinary final examination. Med Educ 1996;30(4):259-65.
9. Blake RL, Hosokawa MC, Riley SL. Student performances on Step 1 and Step 2 of the United States Medical Licensing Examination following implementation of a problem-based learning curriculum. Acad Med 2000;75(1):66-70.
10. Van der Vleuten CPM. The assessment of professional competence: developments, research and practical implications. Adv Health Sci Educ 1996;1:41–67.
11. Wass V, Van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. Lancet 2001;357(9260):945–9.
12. Van der Vleuten C. Validity of final examinations in undergraduate medical training. BMJ 2000; 321(7270):1217–9.
13. American Educational Research Association American Psychological Association, National Council on Measurement in Education. Standards for Educational and Psychological Testing. Washington, DC: American Educational Research Association 1999.
14. Mennin SP, Kalishman S. Student assessment. Acad Med 1998; 73(9 suppl): S46–54.
15. Premadasa IG. A reappraisal of the use of multiple choice questions. Medical Teacher 1993; 15(2-3):237-42.
16. Edward M. Multiple choice questions: their value as an assessment tool. Lippincott Williams & Wilkins, Inc, 2001. Volume 14(6), pp 661-666.
17. Elstein AS. Beyond multiple-choice questions and essays: the need for a new way to assess clinical competence. Acad Med 1993; 68:244-9.
18. Lowe D. Set a multiple choice question (MCQ) Examination. BMJ 1991;302(6779):780-2.
19. Mccoubrie P. Improving the fairness of multiple-choice questions: a literature review. Med Teach 2004;26(8);709–12.
20. Case SM, Swanson DB. Constructing Written Test Questions for the Basic and Clinical Sciences, 3rd ed. Philadelphia: National Board of Medical Examiners, 2001.
21. Anderson J. Multiple-choice questions revisited. Med Teach 2004;26(2):110–3.
22. Hammond EJ, Mclndoe AK, Sansome AJ, Spargo PM. Multiple-choice examinations: adopting an evidence-based approach to exam technique. Anesthesia 1998;53(11):1105-8.
23. The Pakistan College of Physicians and Surgeons. Guidelines for the Development of High Quality Multiple Choice Questions, 2003.
24. Tim Wood T and Cole G. Developing Multiple Choice Questions for the RCPSC Certification Examinations. The Royal College of Physicians and Surgeons of Canada office of Education, September; 2001.
25. Schuwirth LW, Verheggen MM, van der Vleuten CP, Boshuizen HP, Dinant GJ. Do short cases elicit different thinking processes than factual knowledge questions do? Med Edu 2001;35(4):348-56.
26. Ebel RL. Essentials of Educational Measurement, 3rd edn (New Jersey, Prentice-Hall), 1979.
27. Harden RMcG, Brown RA, Biran LA, Dallas Ross WP & Wakeford RE. Multiple choice questions: to guess or not to guess. Medical Education 1976;10;27-32.
28. Thorndike RL. Educational Measurement, 2nd edn, Washington, DC, American Council on Education, 1971.
29. Fleming PR. The profitability in `guessing' in multiple choice question papers. Medical Education 1988;22(6):509-13.
30. Holsgrove G, Elzubeir M. Imprecise terms in UK medical multiple-choice questions: what examiners think they mean. Med Educ 1998;32(4):343-50.
31. Budescu DV, Nevo B. Optimal number of research options: an investigation of the assumption of proportionality. J Educ Meas 1985;22:183–96.
32. Ebel RL. Expected reliability as a function of choices per item. Educ Psychol Meas 1969;29:565–70.
33. Grier JB. The number of alternatives for optimum test reliability. J Educ Meas 1975;12 :109–13.
34. Haladyna TM, Downing SM. How many options is enough for a multiple-choice test item? Educ Psychol Meas 1993; 53:999–1009.
35. Lord FM. Optimal number of choices per item: a comparison of four approaches. J Educ Meas 1977;14:33–8.
36. Veloski JJ, Rabinowitz HK, Robeson MR, Young PR. Patients don't present with five choices: an alternative to multiple-choice tests in assessing physicians' competence. Acad Med 1999; 74(5):539–46.
37. Swanson DB, Holtzman KZ, Clauser BE, Sawhill AJ. Psychometric Characteristics and Response Times for One-Best-Answer Questions in Relation to Number and Source of Options. Acad Med 2005;80(10 suppl): S93-S96
38. Kreiter CD, Ferguson K, Gruppen LD. Evaluating the usefulness of computerized adaptive testing for medical in-course assessment. Acad Med 1999; 74(10):1125-8.
39. Case SM, Swanson DB. Extended matching items: a practical alternative to free response questions. Teach Learn Med 1993;5:107–15.
40. Beullens J, Van Damme B, Jaspaert H, Janssen PJ. Are extended-matching multiple-choice items appropriate for a final test in medical education. Med Teach 2002; 24(4):390–5.
41. Case SM, Swanson DB, Ripkey DR. Comparison of items of five-option and extended-matching formats for the assessment of diagnostic skills. Acad Med 1994; 69(10 suppl): S1-S3.
42. Swanson DB, Case SM. Variation in item difficulty and discrimination by item format on Part I (basic sciences) and Part II (clinical sciences) of U.S. licensing examinations. In: Rothman A, Cohen R (eds). Proceedings of the Sixth Ottawa Conference on Medical Education. Toronto: University of Toronto Bookstore Custom Publishing, 1995: 285–7.
43. Veloski JJ, Rabinowitz HK, Robeson MR, Young PR. Patients don't present with five choices: an alternative to multiple-choice tests in assessing physicians' competence. Acad Med 1999;74(5):539-46.
44. Fenderson BA, Damjanov I, Robeson MR, Veloski JJ, Rubin E. The virtues of extended matching and uncued tests as alternatives to multiple choice questions. Hum Pathol 1997;28(5):526-32.
45. Fowell SL, Bligh JG. Recent developments in assessing medical students. Postgrad Med J 1998;74(867):18-24.
46. Skakun EN, Maguire TO, Cook DA. Strategy choices in multiple-choice items. Acad Med 1994; 69(10 suppl): S7-S9.
47. Van der Vleuten CPM, Newble DI. How can we test clinical reasoning? Lancet 1995;345(8956):1032-4.
48. Case S.M. and Swanson D.B. Constructing written test questions for the basic and clinical sciences. National Board of Medical Examiners, Philadelphia, 1998.

Appendix 1: MCQ Preparation Form

For initial preparation of an MCQ of the single best answer type
- Choose the system and the theme (e.g. fatigue, antibiotics, enzymes, diet or cardiac arrhythmias) and the context (e.g. diagnosis, management, health maintenance or mechanisms of diseases).

Theme _____ Context _____
- Make a list of 6 to 10 options (the provisional option list). These will later form the branches of the question. They must be homogenous (e.g. all diagnoses, antibiotics, blood gas values etc) and each should be short (normally only one or two words).

Provisional option list:

More than 5 are needed – the list must be short and homogenous.

1.		6.	
2.		7.	
3.		8.	
4.		9.	
5.		10.	

- Now, select ONE of the provisional options and mark it with an asterisk.
- Write a vignette (case scenario) that suits the selected option, to form the question stem. This should normally be between two and five sentences in length.

Question stem:

- Next, look at the other options in the provisional list and cross out any that does not suit the correct option well.
- Finally, reduce the option list to 5, containing the correct option plus four others. The chosen option MUST be an indisputably better answer than the other 4 (although the other 4 do not have to be totally wrong).
- Review: Ask a colleague to review the MCQ without the answers. If the colleague has difficulty, modify the option list or scenario as appropriate.

Appendix 2: Preparation of Extended Matching Item (EMI)

- Choose a system or a specialty (e.g. Cardiovascular, Respiratory, Abdomen, Paediatrics or Obstetric Gynaecology etc), a context (e.g. diagnosis, management, health maintenance or mechanisms of diseases) and a theme (e.g. fatigue, antibiotics, enzymes, diet or cardiac arrhythmias).

Specialty (system) _____ Context _____ Theme _____
- Write the lead – in statement:

For each of the following patients (+ with …) – select ..……………………………..
- Make a list of 10-15 options. They must be homogenous (e.g. all diagnoses, antibiotics, blood gas values, aetiology, prognosis etc) and each should be short (normally only one or two words).

Provisional option list:

More than 10 are needed – the list must be short and homogenous.

1.		6.		11.	
2.		7.		12.	
3.		8.		13.	
4.		9.		14.	
5.		10.		15.	

- Now, select three of the provisional options and mark it with an asterisk.
- Write three vignettes (case scenario) that suit the selected options, to form the question stem. This should normally be between two and five sentences in length.

Question stem 1:-
Question stem 2:-
Question stem 3:-

- Validity and discrimination: Look at the other options in the provisional list and cross out:
  - Any option that also suits the written case scenario.
  - Any option that is clearly wrong or very unlikely to be right.
- Finally, reduce the option list to 10, containing the selected options plus 7 others.
- Review: Ask a colleague to review the EMIs without the answers. If the colleague has difficulty, modify the option list or scenario as appropriate.