A Guideline on Developing Effective Multiple Choice Questions and Construction of Single Best Answer format

T BEGUM

Introduction:

The perfect examination would be one in which the student was accurately assessed in his knowledge, comprehension, application, analysis and evaluation of material pertinent to the subject being examined [1, 2]. Although, the essay does determine the candidate's ability to write, assesses powers of logic, original thought and creativity but the use of the essay type question paper as the sole means of assessment has been criticized because of its low reliability and subjective marking [1, 3]. Besides being liable to subjective marking, the area covered by such an examination is very limited and the marking of essays is time consuming.

For several years educationalists of many different disciplines have sought methods of objective testing. In the objective tests the student has to choose the correct response out of one or more alternatives and the marking of the answers is objective [4]. The subjective judgment of the examiner thus plays no part in this form of examination. Multiple choice questions (MCQs) are being increasingly used as objective test for assessment in various fields of education [5, 6, 7].

The characteristics of MCQs are purely objective in scoring and testing, ease of marking, and the ability to test large numbers of candidates with minimal human intervention [8, 9]. Hence, they are cost efficient and feasible. MCQs are time efficient as broad content areas can be tested in a short period of time [1, 4, 5, 8, 9, 10,11]. This allows highly reliable examination.

Disadvantages of MCQs are that they allow for guessing and they are difficult and time consuming to construct [5]. They inhibit students from expressing creativity or demonstrating original thinking. Longer reading time

is required and success depends on suitability of distracters [11].

This article will concentrate on providing an overview of established guidelines for writing effective MCQs so that MCQs can continue to have an important role in assessment and a positive effect on learning.

Question types:

The following are example of some of the MCQ format:

1. **Simple true-false question** is a specialized form of the multiple choices format in which there are only two possible alternatives- either true or false. These questions can be used to measure a student's ability to identify whether statements of fact are accurate or not. It is a very efficient method of testing a wide range of material in a short period of time [12].

2. **Multiple true false (MTF)** or multiple response questions is another format in which candidate choose more than one response from a list of possible answers [12, 13]. There are no restriction on the number of answers to which the correct response is 'true' and the number, to which the correct response is 'false'. It allows a series of questions to be asked relating to a single topic [8, 12, 14].

3. **Single best response or answer (SBA)** - This format consists of a list of possible answers, among which, only one is the "best" and the remaining are inferior but not incorrect [8, 13, 15]. The student is instructed to select a single correct answer or best or most appropriate choice from the group of five possible answers [10, 16, 17, 18, 19, 20].

4. **Matching questions-**It consists of two lists of statements, words or symbols, which have to be matched with one another. Information in each column should be homogenous [20]. These formats are well adopted for assessing student's

Address of Correspondence: Professor Tahmina Begum, FCPS, MD, M MEd, Head, Dept. of Paediatrics, BIRDEM, Dhaka.

understanding of relationship between large amounts of factual information and also application of knowledge [12]. However, it is very difficult to test higher level of cognition with this format [20].

5. **Extended matching question (EMQ)** - as the question is focused on clinical problem, extended matching items aim at assessment of application of knowledge, decision-making or problem solving [12]. The large number of alternatives reduces the effect of cueing. Because the items are relatively short and can be answered quickly, EMQ covers a large knowledge base with a short period of time [12, 13].

7. **Multiple true false completion type-** consists of a stem followed by 4 or 5 true or false statements. Instructions should be clearly given at the beginning [14, 20].

8. **Assertion and reason-** combine elements of true false questions and multiple responses. The question consists of two statements, an assertion and a reason.

The student must first determine whether each statement is true or not. If both are true, the student must next determine whether the reason correctly explains the assertion. There is one option for each possible outcome [12, 21]. This format can be used to explore cause and effect and identify the relationship. Hence, requires a higher level of learning [21].

Among all question types, 3 major formats are being used increasingly, these are:

•   Multiple true false (MTF)

•   Single best answer (SBA)

•   Extended Matching Question (EMQ)

Traditional MTF questions have been considered as most common format historically but increasingly abandoned by many centers over the last decade [22, 23]. MTF format is difficult to write. In order to avoid ambiguity, the writer is pushed to assessing factual recall of isolated fact [21]. Moreover, cueing effect and guessing are major intrinsic drawbacks of this format [9].

In general, Single Best Answer (SBA) format can be invaluable to test application of knowledge, problem solving, judgment and discrimination to a greater extent than other form [9, 15, 18, 21]. These competencies are essential in clinical practice. Furthermore, well-

constructed SBA allows the candidate to demonstrate that they know how rather than simply know and this is a fundamental principle of the assessment of clinical skill [9, 15, 24].

Although probably no easier to write than MTF, the format is more flexible and the chance of random guessing can be reduced [15, 20].

In EMQs, one or more correct response(s) is/are selected from a list of possibilities. EMQs are difficult to construct but more reliable, more discriminating and testing time is reduced [15, 21].

Guidelines for writing MCQs:

Several authors have outlined the elements of good MCQs [7, 25, 26]. Different terms are applied to the different components of MCQs. The *item* is the entire test question, which consists of a *stem,* question or *lead in* and several *options* [5, 12, 13]. Question should focus on an important concept of curriculum [9].

The *stem* consists of a clinical scenario or statement. Compared with MTF questions, in SBA and EMQ, the stem should be long and the options short. It should contain all the relevant information (signs, symptoms, lab test etc.) that is necessary to answer the question.

The possible answers are called *options, alternatives or choices* related to the stem. The correct answer is called the *key* while rest of the options is called the *distracters* or *foils* [3, 5, 11].

•   Writing stems with example:

The stem is usually the first part. It should be brief and clear. Items which use just one or two words in the stem are better avoided. Direct and complete statement is preferable (e.g. which of the following characteristics are the features of . . . .) as incomplete stems lowers the students' correct response rate by 10 – 15% [27, 28].

The stem for MTF may be:

a) A single ward e.g. Digoxin.

b) A statement e.g. contraindications to the use of Digoxin include –

c) Photograph.

d) X-ray of the chest.

e) ECG

f) Diagram, graphs

g) Biochemistry, hematology or other report.

h) A statement from a text

OR A short description of a problem, which may be a case history or clinical scenario ± pathological data ± radiological findings (not more than 50 words) in case of single best answer question (SBA).

- Writing a question or lead in:

It should be clear and short, asking what is the best answer or a similarly phrased question. e.g. 'the most likely diagnosis is …..'

- Writing the key or correct answer/s:

The correct answer is one or the candidate make a single mark on the answer for each question in case of SBA or it may be more than one in MTF question. The key should be clearly correct, should not be longer than other distracters and should not contain any clue [13].

- Writing distracter:

The distracters may not be completely wrong or incorrect or totally farfetched but for the given scenario, less correct [9]. It should differ from key and should not be so close to the correct answer that may confuse the students who know the correct answer. The most challenging aspect of creating MCQs is designing distracters and it is time consuming [1, 5, 17, 29, 30]. The function of the distracters is to distract those students who are uncertain of the answer [11]. The standard of an objective test is probably best assessed in terms of the quality of its distracters. Distracters should be plausible and attractive alternative to the key and should be homogenous or fall into the same category e.g. signs, diagnosis, risk factors, test, treatment etc. They should appear as similar as possible to the correct answer in terms of grammar, length and complexity [5, 12, 13]. The number of possible answers can be 3 to 10 but the reliability is high with 5 options [31]. The list of options should be arranged systematically (alphabetical, chronological, and numerical). There should be sufficient number of distracters to reduce chance of guessing and they should not contain clues. For a distracter to be useful, it should represent a common misconception among students about the correct answers [30].

Points to be remembered during construction of question:
The examiners should know exactly what he/she wants to ask and they have to select the words or sentences that cannot be misunderstand.

During construction of question, following points need to be remembered [4, 5, 9, 12, 13, 15, 32]:

- The stem and options must be clear, concise, unambiguous and correct in grammar and sentence.

- Questions must not contain clues to the correct answer for the student.

- The options should be continuous with the stem e.g.

  Q. Cyanacobalamin (Vitamin B$_{12}$)-

  X Required for folic acid metabolism

  ✓ Is required for folic acid metabolism

- Options in one item should not reveal information that allows the candidate to automatically know the correct answer to another item. This is referred to as 'cueing' when an option in one item provides a hint to the answer for another item.

- Terms such as 'invariable', 'always', 'must', 'all', 'only', and 'never' should be avoided since they imply absolutes and are therefore likely to be false. e.g.

  a) Haemophilia *never* occur in female

  b) Hepatomegaly is *always* found in PEM

- Similarly ambiguous/ imprecise/ too open term like sometimes, possible, may, often, commonly, rarely, usually, can, a few must also be avoided for MCQs to generate valid scores as these terms are difficult to interpret, have different meaning to different person and the answer tends to be true. E.g.

  a) Weight loss is *sometimes* found in thyrotoxicosis

  b) Weight loss is *commonly* found in thyrotoxicosis

  c) Weight loss is found in *a few* patient with thyrotoxicosis

  d) Malarial parasite *often* found in blood film during febrile phase

  e) The *possible* side effect of digoxin are . . .

- The term 'typical' is a useful one for multiple-choice questions. Its meaning implies 'that which is found most commonly'. 'Characteristic' implies a time honored anatomical feature, and 'recognized' an accepted textbook description. e.g. Mental retardation is a *recognized* features of congenital hypothyroidism OR the *characteristic* feature of hypothyroidism is short stature.

- Should not include the phrases *none of the above or all of the ab*ove [5].

- The question like "*which of the following are correct?*" should not be used, as the question is unfocused

- Eponyms, acronyms or abbreviations without some qualification after each term should be avoided.

  E.g. the most likely cause of <u>BPD</u> is . . . .

  Examinees may be unfamiliar with such terms or the terms may have more than one meaning e.g. BPD stands for bi-parietal diameter or broncho-pulmonary dysplasia

- Plausible alternative should be used in options; otherwise it will give a clue to correct answer. e.g

The following vitamin deficiencies are common in children:

  a. Folate   all- vitamin
  b. Riboplavin   e- micronutrient
  c. Vit A
  d. Vit D
  e. Zinc

- Avoid *negative* statement in the stem or option.

  e.g. The following *are not* the feature of Addison's disease

- Avoid *double negative* in the stem or option.

  e.g. The following features are *not unusual* in DKA

- Avoid *double statement* in the stem or option as the student may not be certain as what part is true and what other part is wrong.

  e.g. The *ECG* & *serum cholesterol* level are used to diagnose hypothyroidism.

- Avoid *opposite statement* in the stem or option.

  e.g. SIADH is characterized by:

  i) hyponatremia
  ii) hypernatremia

- Options should be placed in logical order. It is advisable to randomize the order. For example, if the answer is a number, the option should begin with smallest value and proceed to the largest value or largest to smallest. If the options are date, they should list in chronological order. If the options are statement, they should follow *alphabetical* order. If the options are ranges of values, the choice can be independent.

- Common element in the option should be included in the stem. E.g.

The risk factors of pneumonia-

  a) Includes passive smoking
  b) Includes malnutrition
  c) ……………………..

Correct stem is- the risk factors of pneumonia includes-

  a) passive smoking
  b) malnutrition

- Be specific: Inclusion of more information in the question may result in less confusion. e.g. patient's age or sex or habit. What may be appropriate treatment in a 10 month old child, may not be appropriate in a 60 year old person.

- Be careful in the use of figures e.g. In Bangladesh 54% of under five children are malnourished. Actually it is not our intention to test whether the student knows the figure- 54% or 56%. The question can be constructed in a better way. e.g. In Bangladesh, over half of under five children are malnourished or between 50-60% of the under five children are malnourished

- Ensure that the warding of question is precise e.g. the plasma ACTH is high in Addison's disease. The answer would be 'true' if untreated and in primary cause or 'false' if treated or is secondary to pituitary cause. Good question would be 'Plasma ATCH is high in untreated . . . .'

Construction of test items:

Construction of good test items is pre-requisite in preparing effective MCQs [10, 17, 33].

This requires a sound knowledge on the subject and understanding of the objectives of assessment as well as time and skills of writing good test items [17, 34].

If the question is incorrectly or ambiguously worded or is not concerned with appropriate objectives, it will not be reliable.

The steps are- a) developing educational objectives, b) defining levels of learning for each objective and c) writing effective MCQs that test the learning [5].

A direct relationship between instructional objectives and test items must exist. Thus test items should come directly from the objectives [11]. Understanding of and thoughtfully written objectives are critical to the construction of appropriate test questions and in ensuring adequate assessment of intended learner competence [5].

In 1959, Bloom published taxonomy of cognitive domain, which was described as a hierarchy of knowledge [26]. During test construction, the examiners must decide or have guidance on what subject or aspect of the subject is to be tested. Although MCQs are used frequently to test terminology and factual recall, but they can be used to assess higher order of thinking such as application of facts,

interpretation, analysis of relationship, synthesis, problem solving, clinical judgment, reasoning and even attitudes [15]. Inclusion of different level of learning increases the validity of examination [13, 32].

Following examples are given to explain the different level of learning.

Factual knowledge/ Recall:

Q. The risk factors for respiratory distress syndrome (RDS) are:
   a) Asphyxia in term baby
   b) Caesarian section without labour
   c) Intrauterine growth retardation
   d) Maternal diabetes
   e) Prematurity

Understanding:

Q. The mechanism of respiratory distress syndrome in diabetes mellitus is/are:
   a. Delay in development of lung
   b. Decreased stimulation by adrenergic hormone
   c. Decreased production of phosphatidyl- choline
   d. Increased chance of peri-natal asphyxia
   e. Mutation of surfactant protein C gene

Application:

Q. "A baby of Rh –ve mother born at 36 weeks of gestation by C/S. His cord blood bilirubin is 7 mg/dl. The previous 2 sibs died of jaundice in 1st week of life. The steps of management include:
   a) Anti D Ig M to the mother
   b) Anti D Ig G to the mother
   c) Blood transfusion
   d) Immediate exchange transfusion.
   e) Photo therapy

Steps of developing Single Best Answer (SBA) MCQs in short: Example-

Specialty: Paediatrics

System: Haematology

Choose a **theme** (e.g. purpura, gum bleeding, **pallor**, lymphadenopathy, splenomegaly, bone pain etc.)

Choose a **sub-theme** (S/S, risk factors, diagnosis, **investigation,** treatment, complication)

Make a **list of 6-10 homologous options** related to theme & sub-theme (brief, similar length)

Select one option with an asterisk *

Write a stem with case scenario

Write lead in question

Reduce option list to 5 (alphabetical)

Case scenario should include-
• Age, gender
• Site of care (OPD, ER, Ward, ICU)
• Presenting complaints, duration
• Physical finding
• Investigations/X- Ray
• Initial treatment

Homologous option list related to theme (pallor) and sub theme (investigation)
1. Hb%
2. Hb electrophoresis *
3. Coomb's test
4. Bone marrow study
5. Reticulocyte count
6. Serum iron
7. Serum ferritin
8. Osmotic frazility test

Clinical case scenario:
A 5 year old boy attended in OPD with complaints of gradual pallor for 3 years. O/E he was found moderately pale with hepatosplenomegaly. His Hb% was 7 gm% & PBF showed microcytic hypochromic picture with fair number of target cell.

Lead in: What investigation you want to do to confirm the dx?

Option list:
a. Bone marrow study
b. Coomb's test
c. Hb electrophoresis
d. Reticulocyte count   key- c
e. S. ferritin

Another example:
Scenario: A newborn baby delivered by C/S for breech presentation, on routine checkup, his arm was found in a position of adduction and internal rotation with pronation of forearm.

Lead in: Which of the peripheral nerves are injured in this case?

Options:
a) 3rd, 4th and 5th cervical
b) 5th and 6th cervical
c) 7th and 8th cervical
d) 7th , 8th cervical and 1st thoracic
e) 1st and 2nd thoracic

Any TOPIC can form the basis for options in SBA

List of possible themes:

Arteries	Metabolic defects	Physical signs/ symptoms
Muscles	Electrolyte abnormalities	Diagnoses
Nerves	Hemodynamics	Crucial investigations
Cells	Causative agents	Lab studies
Cell components	Pathological process	Initial management steps
Blood components	Cytokines	Management alternatives
Body fluid	Immune disorders	Risk factors
DNA analysis	Vaccines	
Hormones	Toxic agents	
Enzymes	Drugs/ side effects	
Amino acids	Screening tests	

Procedure for constructing the question paper:

Steps to the followed:-

1. Confirm the type and number of objective questions to be used. Forty 5 stem questions can usually be covered in one hour or sixty questions in one and half hours. If problems are given in the stem, than longer time may be necessary.

2. Determine the content of the examination in terms of subject matter, competences to be tested and level of difficulty.

3. Review each item for accuracy and appropriate formatting. Internal review may not reveal all errors. It can be very beneficial to have a colleague read and respond to the MCQs and offer feedback.

4. A question bank may be prepared from where question can be used in future.

A satisfactory bank of questions takes three to five years to build. After this time the questions can be grouped into sections and, whenever an examination paper is required, questions can be chosen at random from each sections. Continuous updating and revision of this material should be undertaken and new material added regularly.

5. During moderation, the following points should be checked-

    a) Wording of the stem & option.

    b) Competences- from recall of facts to higher level of cognitive domain.

    c) Inclusion of different aspects of subject.

    d) Duplication or overlap by using a text matrix.

    e) The instruction to the student, marking scheme or the method of scoring.

6. After examination, examiner should meet to review the questions and the student performance in each item.

Fig: Summery of guidelines for writing effective MCQs [5]

Test items

• Relate directly to instructional objectives

• Test at the same level of learning as the objectives are designed to assess

• Reflect different levels of learning (recall, comprehension, application and problem solving

Stems

• Provide a complete statement

• Include only relevant information

• Contain as much of the test item as possible

• Ask for the correct, not the 'wrong' answer

• Avoid absolute terms, such as always, never, all, or none

• Avoid imprecise terms such as seldom, rarely, occasionally, sometimes, few or many

• Avoid cues, such as may, could or can

• Define eponyms, acronyms or abbreviations when used

Options

- Keep options grammatically consistent with the stem

- Link options to each other (e.g. all diagnoses, tests, treatments)

- Write distracters similar in grammar, length, and complexity

- Write distracters being plausible but clearly incorrect

- Avoid none of the above and all of the above

- Place options in a logical order (e.g. numerical, chronological)

Scoring:

Multiple questions appear to be the easiest of all to mark [32]. A student is either right or wrong and should be given a mark for the correct answer and nothing for the wrong one. This is the simplest and best method of marking [13].

It can be done by computer or by hand if the number of candidates is small. Whatever the scoring system employed, an MCQ paper can place the student in rank order in terms of their ability to answer that particular paper. The real problem of marking MCQ is that of allowing for guessing [31].

Conclusion:

Assessment of learner's knowledge is an important step in educational process. MCQ is the most efficient form of written assessment, being both reliable and valid by broad coverage of content. Improvement of individual's skill in constructing MCQs is crucial. Flawed MCQs interfere with accurate and meaningful interpretation of test scores and negatively affect student pass rate. Therefore, it is important for test developers to be skilled in effective test item writing to evaluate students' competence.

References:

1. Lumley JSP, Craven JL, editors. MCQs in Anatomy. 3rd ed. Newyork: Churchill Livingstone; 2004: p-2-4

2. Crossley J, Humphris G, Jolly B. Assessing health professionals. Medical Education 2002; 36: 800-804

3. Singh T, Natu MV, Singh D, Paul VK, editors. Better Paeditric Education. 1st ed. New Delhi: IAP education center; 1997: p-57-72

4. Newble D, Cannon R, editors. A handbook for medical teacher. 4th edition. Boston: Kluwer Academic Publication 1996; p-140-148

5. Collins J. Education Techniques for life-long learning. Writing multiple-choice questions for continuing medical education activities and self assessment module. Radiographic 2006; 26: 543- 51

6. 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: 593- 96

7. Farley JK. The multiple choice test: Writing the questions. Nurse Educ 1989; 14: 10-12

8. Kehoe J. Basic item analysis for Multiple Choice Tests in practical assessment. Research and evaluation 1995; 4(10): 5-10

9. Mc Coubrie P. Improving the fairness of multiple choice questions: A literature review. Medical Teacher 2004; 26(8): 709-12

10. Epstein RM. Assessment in Medical Education. N Eng J Med 2007; 356: 387-96

11. Gronlund NE. Assessment of student achievement. Boston, Mass: Allyn & Bacon, 1998

12. McBeth RJ, editior. Instructing and Evaluating in Higher Education. Guide book for planning learning outcome. New Jersey: ETP. 1992

13. Schuwirth LWT, Vander Vlenten CPM. Written assessments. In: Dent JA, Harden RM, editors. A practical guide to medical teachers. 2nd ed. Ediburgh: Churchil Livingstone; 2005: p-311-312

14. Amin Z, Eng KH. Editors. Basics in Medical Education. 1st edition. New Jersy: World Scientific; 2003: p-283-98

15. McCoubrie P, McKnight L. Single best answer MCQs: a new format for the FRCR part 2a exam. Clinical Radiology 2008; 63: 506-10

16. Schuwirth LW, van der Vleuten CP. Different written assessment methods: what can be said about their strengths and weaknesses? Med Educ 2004; 38: 974-79

17. Haladyna TM. MC formats. In: Haladyna TM editor. Developing and validating multiple choice test item. 3rd ed. Mahwa, New Jersy: Lawrence Erlbaum Associates; 2004: 67-96

18. Twort C, Mucklow J. How to write high quality questions for a high stakes examinations. Clinical Medicine 2011; 11(3): 227-30

19. Haladyna TM. Downing SM, Rodriguez MC. A review of multiple choice item writing guidelines for classroom assessment. Applied Meas Educ 2002; 15: 309-34

20. Menon PSN. Multiple Choice Question Construction. In: Sood R, editor. Assessment in Medical Education. 1st ed. New Delhi: KL Wig Center for Medical Education & Technology; 1995: p-35-45

21. Case S, Swanson D. Constructing written test questions for the basic and clinical sciences. Philadelphia: National Board of Medical Examiners; 2001

22. Duthie S,Hodges P, Ramsey I. EMQs; a new component of MRCOG part 2 exam. Obstet Gynaecol 2006; 8: 181-5

23. Burton JL. EMQs- how to write and how to answer. Obstet Gynaecol Reprod Med 2007; 17: 25-8

24. Miller GE. The assessment of clinical skills/ competence/ performance. Acad Med 1990; 65: 563-7

25. Braddom CL. A brief guide to writing better test questions. Am J Phys Med Rehabil 1997; 76: 514-16

26. Schultheis NM. Writing cognitive educational objectives and multiple choice test questions. Am J Health Syst Pharm 1998; 55: 2397- 2401

27. Kent TH, Jones JJ, Schmeiser CB. Some rules and guidelines for writing multiple choice test items. Iowa City: University of Iowa College of Medicine and American College Testing Program. 1974: 6

28. Srinivasa DK, Adkoll BV. Multiple choice questions: how to construct and how to evaluate? Indian J Pediatr 1989; 56: 69-74

29. Lowe D. Set a multiple choice question (MCQ) examination. BMJ 1991; 302: 780-2

30. Vayas R, Supe A. Multiple Choice Question: A literature review on the optimal number of options. The National Medical Journal of India2008; 21(3): 130-33

31. Fischer MR, Herrmann S, Kopp V. Answering multiple choice questions in high stakes medical examinations. Med Educ 2005; 39: 890-894

32. Amin Z, Seng CY, Eng KH. Editors. Practical Guide to Medical Student Assessment. 1st ED. New Jersy: World scientific; 2006: p-37-45

33. Walsh K. Advice on writing multipe choice questions. BMJ Careers 2005.

34. Rodriguez MC. Three options are optimal for multiple choice items: A mete analysis of 80 years research. Educ Meas: Issues Pract 2005; 24: 3-13