Observe Shortcomings in framing Multiple Choice Questions for assessing Medical Undergraduates: A Study

Shagufta Wahab, Syed MD Qaseem, Rizwan A Khan

ABSTRACT

Introduction: Multiple choice questions (MCQs) are the most common format of exams for various medical entrance competitive exams in India at both undergraduate and postgraduate levels. This modality of evaluating a student's knowledge may have many shortcomings, which may adversely affect the selection criteria. The aim of this study is to retrospectively analyze the MCQs of Radiodiagnosis in our medical college and suggest modification/changes for better and more standardized MCQs for our medical graduates.

Materials and methods: A total of 20 MCQs set up by teachers of the Department of Radiodiagnosis of our medical college were selected and given to a group of 24 final-year Bachelor of Medicine, Bachelor of Surgery (MBBS) students preparing for postgraduate entrance exams, to answer in 15 minutes. These MCQs were then subjected to detailed item analysis using ease index (EI) and discrimination index (DI).

Results: A detailed analysis of 20 MCQs set up by teachers showed that only 50% of them were phrased correctly. Rest all of them had some shortcomings which required some modification or change.

Conclusion: Multiple choice questions remain the most common format for medical entrance examinations because of its obvious advantages—cheap for evaluating a large group, examiner bias removed, a large part of the course material covered, etc. But setting a perfect MCQ requires considerable practice and sticking to certain rules of setting MCQs, which is a must for proper assessment of our students.

Clinical significance: Setting a perfect MCQ paper is necessary for proper assessment of students and for the benefit of medical education.

Keywords: Discrimination index, Ease index, Medical education, Multiple choice questions.

How to cite this article: Wahab S, Qaseem SMD, Khan RA. Observed Shortcomings in framing Multiple Choice Questions for assessing Medical Undergraduates: A Study. MGM J Med Sci 2017;4(3):125-129.

Source of support: Nil

Conflict of interest: Nil

INTRODUCTION

There are various modes of assessing knowledge, comprehension, and problem-solving capabilities of students like modified essay questions (MEQs) and MCQs. The MEQ is a compromise between the MCQ and the essay. A well-constructed MCQ will be unambiguous, clearly set to a defined standard, and easy to mark (usually automatically). Construction of appropriate MEQs can be difficult\(^1\) and a major criticism of this form of assessment is that MEQs often do little more than test the candidate's ability to recall a list of facts and frustrate the examiner with a large pile of papers to be hand-marked.\(^2\)

Multiple choice questions are a form of assessment in which the students are asked to select a single best answer. Multiple choice testing is a very common format of taking exams and is almost the most common format of exam in postgraduate entrance examinations for medical graduates all over India.

There is an increased need for self-assessment modules that include MCQs and persons with test item writing skills to develop such modules. Although principles of effective test item writing have been documented, violations of these principles are common in medical education. Guidelines for test construction are related to development of educational objectives, defining levels of learning for each objective, and writing effective MCQs that test that learning. Educational objectives should be written in observable, behavioral terms that allow for an accurate assessment of whether the learner has achieved the objectives. Learning occurs at many levels, from simple recall to problem solving. The educational objectives and the MCQs that accompany them should target all levels of learning appropriate for the given content. Characteristics of effective MCQs can be described in terms of the overall item, the stem, and the options. Flawed MCQs interfere with accurate and meaningful interpretation of test scores and negatively affect student pass rates. Therefore, to develop reliable and valid tests, items must be constructed that are free of such flaws.

Multiple choice items consist of a stem and a set of options. The stem asks a question. The options are the possible answers that the student can choose from, with the correct answer called the key and the incorrect answers called distractors.
There are several advantages of multiple choice format of exam, if questions are set properly and it can be a very effective assessment technique. Multiple choice exams require less time for a comprehensive evaluation of a student’s knowledge if they are well distributed over the entire syllabus. They are a cheap format of assessment and save a lot of time if the number of students is large. They also remove various kinds of bias in evaluation as only knowledge about the topic is taken into account and other irrelevant factors are removed while marking the answer sheets.3

However, they are not without disadvantages. Mostly they evaluate lower domains of learning.4 Also, even if the student does not know the exact answer, he/she has a 25% chance of getting the correct answer with only four options. This is also decided by the type of distractors provided, and the probability of getting the correct answer on guessing increases with ineffective distractors, all of the above or none of the above options. Also the MCQs need to be well distributed throughout the syllabus or otherwise the results will be more of a biased reflection of knowledge in selective areas of a subject. Additionally, it is important that questions set should not be ambiguous as it may cause confusion and pose a threat to the validity of the exam.5, 6

Critics of this format suggest other means of examination and evaluation but because of practical difficulties in implementation, the MCQ format of exam remains the most widely used format in entrance to various postgraduate courses. We need to improve our skill of setting MCQs to make it more effective, unbiased, and a more reliable way of assessment as the future of our medical graduates depends on these entrance examinations.

The article provides an overview of established guidelines for writing effective MCQs, a discussion of writing appropriate educational objectives and MCQs that match those objectives, a brief review of item analysis, analyze their adequacy, and suggest modifications and changes if required for better and more standardized MCQs for our medical graduates.

MATERIALS AND METHODS

The study was carried out after obtaining approval from the institutional ethics committee. We collected MCQs from our Department of Radiodiagnosis set by the teachers of the department—four teachers with five questions each. A total of 20 MCQs were selected and given to a group of 24 final year MBBS students preparing for the exams. Each MCQ had four options, one key with rest three being distractors. The time limit was 15 minutes. These MCQs were then subjected to detailed item analysis using EI and DI. They were also analyzed on other points of MCQ evaluation like flaws related to stem and the options (key including distractors) associated with a particular MCQ.

Following was the question paper:

QUESTION PAPER

Time Limit: 15 minutes

1. All of the following are scoring systems for assessing severity of acute pancreatitis except
   (a) Ranson score
   (b) Glasgow score
   (c) Balthazar criteria
   (d) Bismuth classification

2. A young female presents with abdominal pain and ascites. On CT scan, small cirrhotic liver with gross enlargement of caudate lobe and nonvisualization of hepatic veins was seen. Most likely diagnosis is
   (a) Budd–Chiari syndrome
   (b) Alcoholic cirrhosis
   (c) Portal hypertension
   (d) Metabolic liver disease

3. AFP levels are elevated in all except
   (a) HCC of liver
   (b) Papillary CA of thyroid
   (c) Yolk sac tumor of testes
   (d) Pregnancy

4. Neuropathic joint occurs in
   (a) Diabetic neuropathy
   (b) Tabes dorsalis
   (c) Syringomyelia
   (d) All of the above

5. All are true about osteonecrosis except
   (a) Gradual onset
   (b) Reduced activity on scintigraphy
   (c) Self-limiting
   (d) Focal changes on MRI

6. One gray is equal to
   (a) 1 Rad
   (b) 10 Rad
   (c) 100 Rad
   (d) 1000 Rad

7. Definitive method for diagnosing bronchiectasis is
   (a) Chest X-ray
   (b) Bronchography
   (c) MRI
   (d) HRCT

8. Bone density is increased in
   (a) Osteopetrosis
   (b) Idiopathic juvenile osteoporosis
   (c) Osteogenesis imperfecta
   (d) Achondroplasia
9. Earliest diagnosis of osteomyelitis can be made on
   (a) Plain X-ray
   (b) Skeletal scintigraphy
   (c) USG
   (d) CT scan

10. Miliary shadows on chest X-ray is seen in all except
    (a) Tuberculosis
    (b) Sarcoidosis
    (c) COPD
    (d) Metastasis

11. Bilateral hip enlargement can be due to all except
    (a) Sarcoidosis
    (b) Leukemia
    (c) Tuberculosis
    (d) Congenital cyanotic heart disease

12. Kienbock’s disease is due to avascular necrosis of
    (a) Femoral neck
    (b) Lateral cuneiform
    (c) Lunate
    (d) Scaphoid

13. Standard technique for imaging of the breast is
    (a) Digital mammography
    (b) Screen film mammography
    (c) Contrast USG
    (d) MRI

14. In the investigation of ureteric colic, the high accurate
diagnostic modality of choice is
    (a) Spiral CT
    (b) MRI
    (c) USG
    (d) IVU

15. Which of the following is not a radiological sign of
    scurvy?
    (a) Wimberger’s sign
    (b) Frankel’s line
    (c) Pelkan spur
    (d) Looser’s zone

16. Copper beaten skull radiograph is due to
    (a) Craniosynostosis
    (b) Platybasia
    (c) Cerebral tumor
    (d) Raised intracranial pressure

17. Radioactive isotopes used for permanent implant
    brachytherapy is
    (a) Ir-192
    (b) I-125
    (c) Cs-131
    (d) Co-60

18. Snowstorm appearance in USG is seen in
    (a) Vesicular mole
    (b) Fibroid
    (c) Endometriosis
    (d) Adenomyosis

19. HSG should be performed on
    (a) 0–5 days of menstrual cycle
    (b) 5–10 days of menstrual cycle
    (c) 10–15 days of menstrual cycle
    (d) 15–20 days of menstrual cycle

20. Annual dose limit of radiation exposure for general
    public is
    (a) 100 mSv
    (b) 1 mSv
    (c) 20 mSv
    (d) 5 mSv

The students were then evaluated and scored. Top
four scorers formed the group of high scorers (1/6th) and
lowest four formed the group of low scorers (1/6th) for
calculation of EI and DI. The EI and DI were calculated
as follows:

**Ease index:** Tests how easy or tough a question is. It is
calculated by the formula

\[
EI = \frac{h + \frac{1}{n}}{n}
\]

where

\[h = \text{Number of students answering the question correctly in high scorers group}
\]
\[l = \text{Number of students answering the question correctly in low scorers group}
\]
\[n = \text{Total number of students in the two groups}
\]

A question is considered too easy if EI >70% and tough
if EI<30%; 30 to 70% is acceptable.

**Discrimination index**

Differentiates between fraction of high scorers and low
scorers who got that particular question right.

\[
DI = \frac{h - \frac{1}{n}}{n}
\]

where

\[h/n = \text{Fraction of high scorers answering the question correctly.}
\]
\[l/n = \text{Fraction of low scorers answering the question correctly.}
\]

If DI is <0.15 the question has to be discarded and if it
is >0.25 it is recommended.

There was also subjective analysis of the questions.
Finally, the requisite changes were recommended in the
finalized question paper.

**RESULTS**

Following were the results of this study:

- Item 1 resulted in an EI of 50% and DI of 0.5. Both the
indices were in acceptable range.
- Item 2 gave an EI of 62.5% and DI of 0.375. Both the
indices were in acceptable range.
Item 3 had EI of 37.5% and DI of 0.375. However, there was too much use of short forms in the MCQ as AFP, HCC, and CA which is not recommended for setting ideal MCQs.

Item 4 was too easy for the students with EI of 87.5% and DI of 0.125. Both the indices were outside the acceptable range. The question was too easy and all four students from the high scorers group and three out of four students from the low scorers group got it right, thereby making it a bad MCQ for discrimination between high- and low-performing students.

Item 5 had similar problem with very high EI and too low DI.

Item 6 resulted in an EI of 37.5% and DI of 0.5375. Both the indices were in acceptable range.

Item 7 had an EI of 50% and DI of 0.25. Both the indices were good.

Item 8 resulted in an EI of 50% and DI of 0.25. Both the indices were in acceptable range. However, option “b” proved to be an ineffective distractor with not a single student attempting this distractor.

Item 9 had an EI of 37.5% and DI of 0.375. Both the indices were in acceptable range. Option “a” was an ineffective distractor.

Item 10 resulted in an EI of 50% and DI of 0.5. Both the indices were in acceptable range.

Item 11 had an EI of 25% and DI of 0.25. Ease index was too low—less than 30%—meaning it was a tough question, although the DI was in acceptable range.

Item 12 gave an EI of 62.5% and DI of 0.375. Both the indices were acceptable.

Item 13 had an EI of 62.5% and DI of 0.375. The stem of the question was ambiguous on using the term “standard technique,” which does not clearly define the purpose meant.

Item 14 was answered correctly by four students in high scorers group and one student in low scorers group, with EI of 62.5% and DI of 0.375. Both the indices were acceptable.

Item 15 was analyzed to have an EI of 37.5% and DI of 0.375.

The question 16 was totally ambiguous with three out of four options being correct, therefore it was set improperly.

Question 17 was very tough with EI of 12.5% only and DI of 0.125, probably because it was a question more related to radiotherapy and thus beyond the syllabus of radiodiagnosis taught at undergraduate level.

Item 18 gave an EI of 62.5% and DI of 0.375. Both the indices were good.

Item 19 gave an EI of 75% and DI of 0.25. Probably the question was slightly easy but the greater problem was the overlap between options “b” and “c” as the correct answer seems to lie between the two options and should have been better framed.

Item 20 resulted in an EI of 62.5% and DI of 0.375. Both the indices were acceptable.

**DISCUSSION**

The MCQ is the most common type of written test item used in undergraduate, graduate, and postgraduate medical education. MCQs can be used to assess a broad range of learner knowledge in a short period of time. Because a large number of MCQs can be developed for a given content area, which provides a broad coverage of concepts that can be tested consistently, the MCQ format allows for test reliability. If MCQs are drawn from a representative sample of content areas that constitute predetermined learning outcomes, they allow for a high degree of test validity. Critics of MCQs argue that higher-level learning cannot be tested with MCQs. However, this criticism is more often attributed to flaws in the construction of the test items rather than to their inherent weakness. Appropriately constructed MCQs result in objective testing that can measure knowledge, comprehension, application, and analysis. Disadvantages of MCQs are that they test recognition (choosing an answer) rather than recall (constructing an answer), they allow for guessing, and they are difficult and time-consuming to construct.

Analysis of this study revealed what all shortcomings can occur in MCQs if not framed correctly. For example, we found that:

- Questions may be too easy (MCQs 4 and 5), which may make discrimination between performers and nonperformers difficult.
- While one or two tough questions may be included in the MCQs, they should not be too tough, otherwise DI may become insignificant (e.g., MCQ 17 with DI 0.125).
- Ineffective distractors (like option “b” in MCQ 8 and option “a” in MCQ 9) should not be given as option, because these are likely to be attempted by a small number of students.
- Abbreviations may be used inadvertently in MCQs which may not be understood by students (like HCC for hepatocellular carcinoma in option “a” of MCQ 3).
- It is better to use full form of words.
- There should be no ambiguity in questions as in MCQ 13 (term “standard technique”) and MCQ 16 (3 out of 4 options correct).
- Options given should not be too close or overlapping as in MCQ 19 (Option “c” 10–15 days, option “d” 15–20 days).
- In the present study, it is apparent that only lower domains of learning are being tested. We should test higher levels of learning as well, e.g., by giving real clinical scenarios.
• Stems of MCQs should be positive and not negative. We should ask for “correct” answer rather than “wrong” answer. Absolute and imprecise terms (like always, never, all, seldom, rarely) should be avoided.

Palmer and Devitt\(^8\) compared overall performances of MCQs and MEQs for their abilities to test higher cognitive skills. They found MCQs better than MEQs at addressing higher levels of cognitive skills. Collins\(^9\) suggested that test items in MCQs should relate directly to instructional objectives and should reflect different levels of learning (recall, comprehension and application, problem solving).

**CONCLUSION**

All in all, although the relatively small sample of this study cautions against overinterpretation, based on this study, we suggest that a well-constructed MCQ meets many of the educational requirements and suggest that this format be considered and utilized seriously when assessing students. Benefits of removing examiner bias with automated marking and potentially high reliability at low cost make MCQs a viable option for assessment in clinical medicine. This mode of examination also covers a large part of the course material at the same time. Setting a perfect MCQ requires considerable practice and sticking to certain rules.

**Clinical Significance**

Setting a perfect MCQ paper is a must for proper assessment of students and for benefiting medical education. This mode of evaluation, if done perfectly, will offer various advantages for both the examiner and the examinee, which may be evaluated in future studies.

**REFERENCES**