ITEM ANALYSIS OF MCQS’ - MYTHS AND REALITIES WHEN APPLYING THEM AS AN ASSESSMENT TOOL FOR MEDICAL STUDENTS

Priya S. Patil¹, Manisha R. Dhabale ², Nitin R. Mudiraj³

¹Assistant Professor, Department of Anatomy, Bharati Vidyapeeth Deemed University Medical College and Hospital, Sangli, MS, India; ²Associate Professor, Department of Anatomy, Bharati Vidyapeeth Deemed University Medical College and Hospital, Sangli, MS, India; ³Professor, Department of Anatomy, Bharati Vidyapeeth Deemed University Medical College and Hospital, Sangli, MS, India.

ABSTRACT

Multiple choice questions [MCQs’] form one of the tools for evaluation of medical graduates. The widespread use of MCQs’ raises the question of their validity and authenticity. The present study was focused on MCQ validation with an attempt to highlight the importance of item analysis and discuss the myths and realities of MCQs’.

Aim: The aim was to perform Item analysis while probing into myths and realities of MCQ as an assessment tool for first year medical students. A total of hundred students underwent MCQ test as part of their first periodic assessment.

Methodology: A total of hundred students underwent MCQ test as part of their first periodic assessment. A pre-validated MCQ test was given to students and a post-validation was done through item analysis. The indices were calculated using Microsoft excel and various indices compared.

Results: Inspite of using a well-designed MCQ test validated by senior faculty and colleagues, after item analysis the difficulty index of 65 % items and discrimination index of 75 % items was in the acceptable range. The distractor efficacy showed that 9 out of 120 distractors needed revision.

Conclusions: Item analysis must be routinely implemented to validate MCQs’. The myths like shortfalls of MCQ framing, only four options in clinical setting or cueing effect are duly considered and the realities like strict and elaborate protocol for MCQ construction, stimulating critical and lateral thinking of students should be implemented. An integrated approach can help to achieve an ideal assessment tool for the benefit of students.

Key Words: MCQ validation, Difficulty index, Discrimination index

INTRODUCTION

Evaluation is an important tool that guides student learning. Multiple choice questions form one of the tools for evaluation and are currently being widely used for assessment of medical graduates.¹ Multiple choice questions are particularly useful to assess a large body of material for many students. A multiple-choice question (MCQ) consists of a stem with a question line at its end or underneath it, followed by a number of options. One of the options is correct or best response known as the key, while the others are described as distractors.³ Constructing a valid and effective MCQ is difficult task and needs a different approach. The methodical development of MCQ banks can be an asset which can lead to reliable assessment of students.² Current widespread use of MCQs’ raises the issue regarding their validity and authenticity.³ Review of literature highlights different views regarding use of MCQ or combined approach of MCQ and free response essay type questions for better assessment of medical graduates. Considering the above issues the present study was focused on MCQ validation through item analysis. The importance of item analysis and the myths and realities concerned with MCQ as an assessment tool for medical students are discussed further.

MATERIAL AND METHODS

The study was conducted in the Department of Anatomy as a part of First Periodic Examination. Hundred First-year MBBS students were voluntarily involved in the study. They
were given the MCQ test comprising of forty [40] questions with single best response. The MCQs’ were constructed to assess various levels of knowledge according to Bloom’s Taxonomy. Pre-validation of the MCQs’ was done by the Head of Department and other colleagues in the department. There was no negative marking and the time allotted was forty minutes. Evaluation was done out of forty marks and students were grouped as low, moderate and high achievers.

Post validation of the MCQs’ was done by item analysis. The scores of all the students were arranged in such a way that there was order of merit. The students were divided into three groups according to their marks. Those students who belonged to the first group were considered as high achievers H \( [n=30] \) and those in the third group as low achievers L \( [n=30]\). Each item was analyzed for the following:

1. **Difficulty Index or Facility value or p value** was calculated using the formula
   \[ p = \frac{H + L}{N} \times 100 \]
   Where H= number of students answering the item correctly in the high achieving group
   L= number of students answering the item correctly in the low achieving group
   N= Total number of students in the two groups (including non-responders)

   Items having p value between 30 – 70% are considered as acceptable, among which items with p value between 50-60% are ideal while items with p value less than 30% (too difficult) and more than 70% (too easy) are not acceptable and need modification.1

2. **Discrimination index (DI) or d value** was calculated using the formula
   \[ d = \frac{H - L}{2N} \]
   Where the symbols H, L and N represent the same values as mentioned above.

   The DI or d value is a measure of the item to discriminate between students of higher and lower abilities and ranges between 0 and 1. In general, d values between 0.20 and 0.35 is considered as good. Items having DI values more than 0.35 are considered to be excellent and those with less than 0.20 are considered poor.

3. **Distractor Effectiveness (DE)** or Functionality of the distractors was calculated by observing how many students used the distractor in an item. If less than 5% of students used the option other than the key then that distractor is said to be a non-functional distractor [NFD]. On the basis of number of NFDs in an item, DE ranges from 0 to 100%. If an item contains three or two or one or nil NFDs then DE would be 0, 33.3%, 66.6% and 100% respectively.6

## RESULTS

1. **Difficulty Index of items analyzed**
   After individually analyzing the items it was observed that the difficulty index (p value) of 26 items that means 65% were acceptable. 11 items were too easy and 3 were too difficult hence 14 items that means 35% could be used after modification. [Table 1]

2. **Discrimination index of items analyzed**
   The discrimination index of 30 items [75%] was acceptable but 10 items [25%] need revision due to poor index. [Table 2]

3. **Distractor efficacy [DE] based on number of NFDs’**
   As per tables 3, 4 and Graph 1, distractor efficacy was 100% in 24(60%) items and was 66.66% in 12(30%) items. So these questions can be added to the MCQ bank. In 3 items distractor efficacy was 33.33% hence their distractors should be modified and in 1 item having distractor efficacy 0%, its distractors should be either changed or the question should be discarded. Hence out of 120 distractors 9 need revision.

   It was observed that higher the Difficulty Index, lower was the difficulty of the question. The Difficulty Index and Discrimination Index were often reciprocally related. Questions having a high p value (easier questions) discriminate poorly; while those with a low p value were good discriminators.

## DISCUSSION

Good teaching is more a giving of right questions than a giving of right answers - Joseph Alberts

The selection of an appropriate assessment method for students’ performances remains a daunting task for the teachers. In an attempt to change and upgrade the existing assessment tools MCQs’ are introduced at various levels and are used rampantly for regular assessments as well as important career deciding entrance examinations. A large number of educational research studies have demonstrated poor validity of MCQ’s. Barrows and Tamblyn in their book on Problem Based Learning mention, multiple choice and true/false questions under the heading “reliable evaluation tools with questionable validity”. ‘Item analysis is truly the assessment of the assessment tool itself.’ Post examination analysis of the MCQs helps to assess the quality of individual test items and test as a whole. Poor items can be modified or removed from the question bank. It also helps the teachers to identify the subject content which lacks understanding and need greater emphasis and clarity, by improving or changing the method-
ology of teaching. In spite of validating the MCQ items by seniors and colleagues in the current study post validation systematic item analysis showed lacunae. The difficulty index was acceptable in only 65% items while the discrimination index was acceptable in 75% items. The distractor effectiveness also reflected need for revision in 9 out of 120 total distractors. Similar results were seen in different studies conducted for item analysis of MCQ. They reveal that though item analysis was traditionally considered as tedious and time consuming task, the use of software applications such as Microsoft Excel has made it relatively easy and effective way to calculate various indices and test the validity of the items. Item analysis thus gives an insight and opportunity to revise and improve the MCQ and develop a valid question bank.

MCQ- Myths and Realities

There are many opinions and views of teachers as well as students regarding MCQ assessment and assessment in the form of short essay questions. This has led to the following discussion on the myths and realities about MCQs.

- MCQ as guess items assess only factual knowledge: One of the myth we frequently come across regarding MCQs is that they are multiple guess items which are useful to assess only factual recall of knowledge and higher orders of cognitive skills cannot be assessed. However the reality is MCQs can be framed and well validated in order to assess higher domains of Bloom’s taxonomy. Developing a good bank of valid and reliable multiple-choicetest items targeting higher cognitive abilities while conforming to item construction guidelines presents a big challenge to thetem developer. Yet in reality a dilemma remains during assessment because MCQs give four options to the students. Hence we cannot know exactly whether a particular question was answered by rapid retrieval of overlearned content or by purposeful relating of principles.

- The MCQs’ challenge - Another widely prevalent myth regarding use of MCQ as an assessment tool is that MCQs are easy to frame and less time consuming. In reality it is a challenge for the teachers to construct ideal MCQs’ such that all options present as plausible answers. The distractors should be such that they attract students who do not know the answer while who know the right answer ignore them.

- Cueing effect - MCQs’ are recently used in most of the entrance examinations at various levels. This has probably led to the myth that they are the best and fast method for assessment in competitive examinations. Multiple studies show that if same group of students are assessed by MCQ and free response tests all students scored higher in objective assessment. The cueing effect of readymade options available was thought to be the major factor.

- Integrated approach- If we consider medical professional training it is mandatory that more emphasis be given to academic excellence and clinical competency. The medical students should be trained to evolve as problem solvers and not wait for situations with four options. Moreover there are certain clinical scenarios where there is one and only one diagnosis or some other conditions where there are more than four differential diagnoses which cannot be dealt with in an MCQ. The above dilemmas can perhaps be solved by introducing MCQs as a part of student assessment in undergraduate examinations more so in pre-clinical subjects. Some of the advantages in doing so are that the MCQs can be assessed rapidly with the aid of computers and softwares so can be employed when large number of students are to be assessed along with fast results and less error. They can provide highly structured and clear tasks with easy, objective and reliable scoring where legibility of handwriting doesn’t matter. Such data can be compared from class to class and year to year. Such assessment with MCQs can also be applied to certain clinically relevant topics like Paediatrics. Later in the medical coursewhile using MCQ as an assessment tool for students they should undergo strict protocol for their framing, validation and selection. They should be overweighted by free response questions and newer ways of assessment to enhance critical thinking of students like case based learning, early clinical exposure, clinical scenarios and problem solving can be introduced. The MCQ test in entrance examinations can be supplemented with well-designed OSCE modules to assess students’ performances.

In studies conducted by various authors it was seen that most of the students had a consistent performance in both objective and theory assessment. Interestingly it was also noticed that in some cases there was no correlation between MCQ and long essay scores for either the more competent students or the students who received failing grades. Thus, students’ grades determined by an examination format that includes both testing modalities may be different than the grades obtained by using only one of the modalities.

**CONCLUSION**

Item analysis enhances the quality of MCQs which can help create an authentic and validated MCQ bank which can also assess higher orders of cognitive skills. The myths like shortfalls of MCQ framing, only four options in clinical setting or cueing effect must be considered during the use of MCQ as an assessment tool. The realities like strict and
Patil et al.: Item analysis of MCQs’ - Myths and Realities when applying them as an Assessment Tool...

elaborate protocol for MCQ construction, integration with well-designed free-response and short-essay-type questions, stimulating critical and lateral thinking of students should be implemented. Such a perspective can help to achieve an ideal assessment tool for the benefit of students at large.

Table 1: Difficulty Index of items analyzed

<table>
<thead>
<tr>
<th>DI Range</th>
<th>Number of items</th>
<th>Interpretation</th>
<th>Action taken</th>
</tr>
</thead>
<tbody>
<tr>
<td>30 to 70 %</td>
<td>26 [65%]</td>
<td>Acceptable</td>
<td>In MCQ bank</td>
</tr>
<tr>
<td>&gt;70 %</td>
<td>11 [27.5%]</td>
<td>Too easy</td>
<td>Revise</td>
</tr>
<tr>
<td>&lt;30 %</td>
<td>3 [7.5%]</td>
<td>Too difficult</td>
<td>Revise</td>
</tr>
</tbody>
</table>

Table 2: Discrimination index of items analyzed

<table>
<thead>
<tr>
<th>Range</th>
<th>Number of items</th>
<th>Interpretation</th>
<th>Action taken</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2 to 0.35</td>
<td>13</td>
<td>Good</td>
<td>In MCQ bank</td>
</tr>
<tr>
<td>&gt;0.35</td>
<td>17</td>
<td>Excellent</td>
<td>In MCQ bank</td>
</tr>
<tr>
<td>&lt;0.2</td>
<td>10</td>
<td>Poor</td>
<td>Revise</td>
</tr>
</tbody>
</table>

Table 3: Distractor efficacy

<table>
<thead>
<tr>
<th>Distractor type</th>
<th>Number of items</th>
<th>Action taken</th>
</tr>
</thead>
<tbody>
<tr>
<td>Functional</td>
<td>99</td>
<td>In MCQ bank</td>
</tr>
<tr>
<td>Non-functional NFD</td>
<td>21</td>
<td>Revise</td>
</tr>
</tbody>
</table>

Table 4: Distractor efficacy [DE] based on number of NFDs’

<table>
<thead>
<tr>
<th>NFD</th>
<th>Number of items</th>
<th>DE</th>
<th>Action taken</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>24</td>
<td>100%</td>
<td>In MCQ bank</td>
</tr>
<tr>
<td>1</td>
<td>12</td>
<td>66.66%</td>
<td>In MCQ bank</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>33.33%</td>
<td>Revise</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>Revise / Discard</td>
</tr>
</tbody>
</table>

Graph 1: Distractor efficacy [DE] and number of items

ACKNOWLEDGEMENTS

Authors acknowledge the immense help received from the scholars whose articles are cited and included in references of this manuscript. The authors are also grateful to authors/editors/publishers of all those articles, journals and books from where the literature for this article has been reviewed and discussed. I sincerely thank the institute as well as all the students, colleagues, office clerk and non-teaching staff for their co-operation during this study.

Sources of funding: Nil

Conflict of interest: Nil

REFERENCES


