Original article

Item analysis of Multiple Choice Questions in Physiology examination.

Dr. Vrunda Kolte

Department of Physiology
Name of the Institute/college: Lokmanya Tilak Municipal Medical College & General Hospital
Corresponding author : Dr. Vrunda Kolte

Abstract

Introduction: Multiple choice questions (MCQs) /Items are frequently used to assess students in different educational streams for their objectivity and wide reach of coverage in less time. However the items to be used must be of good quality. The present study was conducted in a medical college of Mumbai with the Objective of evaluation of items to develop a pool of valid items & to update question bank for designing question paper as per the need of assessment.

Methods: Total 40 items from physiology preliminary examination of 100 students of 1st year M.B.B.S. were analysed. Each item was analysed for Difficulty index, Discrimination index and Distracter effectiveness.

Results: After statistical analysis of data, it was observed that Difficulty index i.e. ‘p’ value of analysed MCQs ranged from 6.25% (lowest) to 90.6% (Highest) & Discriminative index i.e. ‘d’ value ranged from 0 (lowest) to 0.63 (Highest). Total 65% items were in acceptable range of difficulty level (‘p’ value 30 – 70%) & 10% items were very difficult which later discussed with students. Discrimination index of 60% items was excellent (d value>0.35). No item had Negative discriminative power. About 47.5% items had 100% Distracter Efficiency (DE) whereas 7.5% items had 0% DE.

Conclusion: Item analysis definitely helps to update & strengthen MCQ bank. It helps in question paper setting as per the need of assessment. It improves Teaching – Learning outcomes.

Key words - item analysis, difficulty index, discrimination index, distracter effectiveness

Introduction

Evaluation is an important component of a teaching-learning curriculum. It is thus important to realize that while selecting the tool for evaluation, the qualities of the tool in terms of validity, reliability and objectivity need to be considered.\(^1\,\) Multiple choice questions (MCQS) are used as an objective and reliable tool to evaluate learning performance of students. It is also a preferred tool for selection of students for a given course. Properly constructed MCQs can assess higher cognitive processing of Bloom’s taxonomy such as interpretation, synthesis and application of knowledge instead of just testing recall of isolated facts.\(^2,\,3\) Designing good MCQs is a complex, challenging and time consuming process. Studies have shown that it is five times faster to revise items that didn’t work, using item analysis, than trying to replace it with a completely new question. New item which would just have new problems. Having constructed and assessed, MCQs need to be tested for the standard or quality. Item analysis examines the student responses to individual test items (MCQs) to assess the quality of those items and test as a whole. \(^4\) It is a valuable yet relatively simple procedure performed after the examination that provides information regarding the reliability and validity of a test. \(^5\)
Materials and methods

The present study was conducted in the Department of Physiology, Lokmanya Tilak Municipal Medical College & General Hospital, Sion, Mumbai. Institutional Ethical Committee clearance was obtained.

Total 40 items of single best response type, from Physiology preliminary examination of one hundred 1st year MBBS students, were analysed. There was no negative marking and the time allotted was one hour. Pre-validation of the paper was done by the experts. Evaluation was done out of forty marks and 50% score was the passing mark. Post validation of the paper was done by item analysis. The scores of all the students were arranged in order of merit. The upper one third students were considered as high achievers and lower third as low achievers.

High achiever group (n= 33) - Starting from the highest rank, 1/3rd of the papers with high scores were selected.

Low achiever group (n=33) - lower 1/3rd of the papers with low scores were selected.

Middle third (n=34) were set aside.

Each item was analysed for:

1. **Difficulty index**: The percentage of total number of students from both the groups (High Achievers & Low Achievers) opting for key (i.e. answering correctly) represents the difficulty index (denoted as ‘p’)

   Difficulty Index (Dif I) 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 achievers group
   - \( L \) = number of students answering the item correctly in the low achievers group
   - \( N \) = Total number of students in the two groups (including non-responders)

   In general, Items with difficulty index less than 30% are considered as difficult. If an item has a ‘p’ value between 30-70% it is considered as acceptable. Items with difficulty index greater than 70% are considered as easy.

2. **Discrimination index** denoted as ‘d’ measures the ability of an item to discriminate between students.

   To calculate this index the numbers of responders to the ‘key’ was taken into account. Here the difference between the two groups was found out. The larger the difference between high achievers and low achievers, the greater will be the discrimination power of an item.

   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 Discrimination index ranges from -1 to +1 .An index value of +1 means the item has maximum discriminative power. An item having a discrimination index greater than 0.35 is considered as to have excellent discriminative power. An item having a discrimination index between 0.2 and 0.35 has acceptable discriminative power. An item having discrimination index ‘0’ can not discriminate between two (H&L) groups. An item having Negative discrimination index ranging from -1 to 0 has poor discriminative power.

3. **Distracter Effectiveness (DE) or Functionality**

   An item contains a stem and four options including one correct (key) and three incorrect (distracter) alternatives. Non Functional Distracter (NFD) in an item is the option, other than the key selected by less than 5% of students and functional or effective
distractor is the option selected by 5% or more students. 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.

Results:

Table No: 1

Categories of Items according to Difficulty index

<table>
<thead>
<tr>
<th>Difficult Index ‘P’</th>
<th>Category of Item</th>
<th>Total Number of Items/40</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;30%</td>
<td>Difficult</td>
<td>4</td>
</tr>
<tr>
<td>30-70%</td>
<td>Acceptable</td>
<td>26</td>
</tr>
<tr>
<td>&gt;70%</td>
<td>Easy</td>
<td>10</td>
</tr>
</tbody>
</table>

Table No : 2

<table>
<thead>
<tr>
<th>Discrimination Index (d)</th>
<th>Discriminative Power</th>
<th>Total Number of Items/40</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥ 0.35</td>
<td>Excellent</td>
<td>24</td>
</tr>
<tr>
<td>0.2 to 0.34</td>
<td>Good</td>
<td>7</td>
</tr>
<tr>
<td>&gt;0 to 0.2</td>
<td>Acceptable</td>
<td>8</td>
</tr>
<tr>
<td>0</td>
<td>Can not Discriminate</td>
<td>1</td>
</tr>
<tr>
<td>&lt; 0 (0 to -1)</td>
<td>Poor</td>
<td>0</td>
</tr>
</tbody>
</table>

With Mean ± SD = 57.92 ± 19.58
Table No : 3

<table>
<thead>
<tr>
<th>Total Number of Items = 40</th>
</tr>
</thead>
<tbody>
<tr>
<td>Items with 0 NFD</td>
</tr>
<tr>
<td>DE = 100%</td>
</tr>
<tr>
<td>19 = 47.5%</td>
</tr>
</tbody>
</table>

Thus, Distracter efficiency of items ranged between 0 – 100% with Mean ± SD = 75.55 ± 29.33

**Discussion**

Previous studies have proposed the mean of Dif I as 39.4± 21.4% [6], 52.53±20.59. [7] Karelia B, showed a range of mean ±SD between 47.17±19.77 to 58.08±19.33 in a study conducted over a period of five years [8]. They also showed 61% items in acceptable range (p 30-70%), 24 % items (p>70%) and 15 % items (p< 30%). Other studies showed that 62% items had p value (30-70%), 23 % were too easy (p >70%) and 15% were too difficult (p<30%). [7]

Patel KA and Mahajan NR showed 80% of items in the acceptable range (p 30-70%) and 20% in the unacceptable range (p >70% & <30%). [9]

My findings corresponded with the previous studies having a mean of Dif I as 57.92 ± 19.58. The p value of 26 (65%) items was in acceptable range (30 – 70%). 10(25%) items were easy with p value > 70% & 4(10%) items were difficult with p value < 30%. On reviewing difficult questions, it was observed that these items were from ‘Desirable to know’ or ‘Nice to know’ portion of syllabus. One of the item was related to common misconceptions and one item which was placed at the very end of the paper, had confusing distracters. Difficult questions were discussed with students which helped them to clear their doubts. On reviewing easy items, it was noticed that many of them were from ‘Must know’ portion of syllabus. Those topics were emphasized during theory lectures and were taught in greater detail. It is advisable to place easy questions also in test paper to boost up confidence of all types of students. Similarly difficult questions can be retained and used to select toppers. As the Dif. I moves towards high or low from 50%, the discriminating index becomes low. [10] Si - Mui et al showed that MCQ items with good discriminating potential tend to be moderately difficult items. [11] It has been seen that the relationship between Dif I and DI is not linear, but predicted as dome shaped [7,8]. For easy items, discrimination may be poor. This is because both, high and low achievers can answer correctly. Whether to retain such item depends on its relevance. If this item is testing a knowledge area, which all the students ‘must know’ then such an item should be stored and used in examination or classroom teaching to find out whether 100% learning has taken
Items with very high discrimination index are given in Entrance exams, in which main purpose of test is to select, only few good students from many.

There are instances when the value of DI can be less than 0 (negative DI), which simply means that the students of lower ability answer more correctly than those with higher ability. This is probably due to complex nature of item, making it possible for students of lower ability to select correct response by guess without any real understanding, while a good student suspicious of any easy question, takes a harder path to solve and ends up to be less successful.[6] Earlier studies have revealed 40% items with DI >0.35, 42% with DI between 0.2 and 0.34 and 18% with DI < 0.20.[9] Another study showed 29% items with DI>0.4%, 46% items with DI between 0.2-0.39 and 21% items with DI < 0.19.[12]

In the present study, the mean of DI was 0.33 ± 0.15. Total 24 (60%) items had Excellent discriminative power. 7 (17.5%) items had Good discriminative power. Total 8 (20%) items had Acceptable discriminative power whereas 1 (2.5%) item showed poor discrimination. 0% of total items had Negative discriminative power. Some studies have shown negative DI in 20%[6] and 4%[13] items. Probable explanation was wrong key, ambiguous framing of questions or generalized poor preparation of students.[6] Items with negative DI decrease the validity of the test and should be removed from the question bank.

Most difficult task in formatting good quality MCQs is writing appropriate options to the correct answer. A distractor analysis gives an opportunity to study the responses made by students on each alternative of the item. NFDs should be removed from the item or be replaced with a more plausible option.[14] In a study conducted on 514 items and 1542 distractors, 35.1% were NFDs, 52.2% were functional distracters and 10.2% were not chosen by any student.[15] Another review of functioning distracters in 477 items on four MCQ assessments showed 38% items had NFDs and items with three functional distracters ranged from only 1.1 to 8.4%. [16]

Mehta G et al have shown, in their study, with fifty MCQs, having 150 distracters, 53 (35.33%) were found to be NFDs, 28 (18.66%) were functional distracters and 69 (46.01%) distracters had nil response. The number of items having NFDs was found to be 33 (66%). On the basis of number of NFDs, items with DE 66.6% were 18 (54.4%), items with DE 33.3% were 9 (27.27%) and items with DE as 0 were 6 (18.18%). The remaining 17 items with three functional distracters had DE as 100%.[17]

Gajjar et al have shown that, in a total of 150 distracters, 133 (89.6%) were functional distracters, and 17 (11.4%) were NFDs. Items with NFDs were 15 (30%) out of which 13 items had DE of 66.6% and 2 items had DE of 33.33%. [6] Our study shows that, in a total of 120 distracters, 91 (75.8%) were functional distracters, and 29 (24.16%) were NFDs. Items with NFDs were 21 (52.5%) out of which 16 (40%) items had DE of 66.6% and 2 (5%) items had DE of 33.33% and items with DE as 0 were 3 (7.5%). The remaining 19 (47.5%) items with three functional distracters had DE as 100%. Students’ performance depends on how distracters are designed. Analysis of the distracters, identifies their errors, so that they may be revised, replaced or removed.[18]

It was observed that in present study question paper, items were included from all levels of difficulty. Easy or acceptable question were placed in the beginning and difficult questions interspersed in between, (60%) of total items had Excellent
discriminative power and no item showing negative discrimination. 47.5% items had DE as 100%. It was well designed question paper. Items having NFDs were discussed with faculty experts and required modifications were done to update question bank. With item analysis, teachers get an insight into how well a particular topic has been understood by the students and which topics need more emphasis. Based on this teachers can modify their teaching either with respect to content or method. Discussion on results of item analysis with students helps them to learn in a better way. It is a revision of the subject they have learnt already, along with clarification of aspects which they have not understood.

**Conclusion**

This study inferred that analyzed and revised test items strengthen and update the MCQ bank.

Item analysis and storage of MCQs with their indices provides opportunity for an examiner to select MCQs of appropriate difficulty level as per the need of assessment and decide their placement in the question paper. Discussion about results of item analysis with faculty as well as with students helps in improving learning outcome. In long term, if we do item analysis over ten years, this data will be a source of great item bank. Also it can assess the effect of any modification in the teaching methodologies.

**References:**

15. Tarrant M, WareJ, Mohammed AM. An assessment of functioning and nonfunctioning distracters in multiple choice questions: a descriptive analysis. BMC Medical Education 2009 ;9:40