RESEARCH ARTICLE

Item analysis of multiple-choice questions in anatomy for first year MBBS

Chandrika Teli¹, Nilesh Kate²

¹Department of Anatomy, ESIC Medical College, Kalaburagi, Karnataka, India, ²Department of Physiology, ESIC Medical College, Kalaburagi, Karnataka, India

Correspondence to: Chandrika Teli, E-mail: drchandrikaesic@gmail.com

Received: January 04, 2022; Accepted: January 27, 2022

ABSTRACT

Background: Multiple choice questions (MCQs) are preferred tools of assessment because of objectivity, ease of scoring and how each MCQ functions as an item can be understood by Item analysis. Aims and Objectives: The aim of the study was (i) to carry out item analysis of MCQs used in formative assessment to know the validity and (ii) to carry out a post validation item analysis of MCQ's of 1st MBBS students in anatomy to use results for further actions. Materials and Methods: 45 MCQs were administered to 112 students of 1st M.B.B.S. as a formative assessment. Difficulty index and Discrimination index were calculated. Results: Mean difficulty index was 56.67 ± 22.09, and mean discrimination index was 0.35 ± 0.23. Distribution of easy, moderate, and difficult MCQ was 20, 67, and 13%, respectively. About 20% of MCQs were poor, 20% with acceptable discriminating index, 27% had good, and 33% MCQ were of excellent discrimination index. No item was negatively discriminating and all distractor were functional. Very easy and very difficult items had poor discriminating index. Conclusion: Most of items had moderate difficulty and good to excellent discrimination. Too easy and too difficult items showcased poor discrimination, no negatively discriminating item and absence of non-functional distractor suggest good quality framing of the MCQs.

KEY WORDS: Item Analysis; Facility Index; Discrimination Index; Single Best Response Multiple Choice Question

INTRODUCTION

Multiple choice question (MCQ) is most preferred tools for written assessment format. Wide content coverage, ease of scoring, and high objectivity and reliability make MCQs-good assessment tool.[1] Properly constructed MCQ can be used with advantage for testing high-order thinking skills.[1]

How each MCQ functions as level of difficulty and in identifying the spread of high and low performers is decided by item analysis[2] item analysis results can give an insight about construct, interpretation, and place of MCQ in question banks. Faulty items can be re-framed, corrected or eliminated for further use. It also serves to provide feedback to teachers about the remedial measures/corrective measures for any content taught, understanding abilities of students, and to plan teaching activities accordingly.[2-4]

“Single best response” type with 4 options is recommended by most of universities for assessment. Recently, the RGUHS University has advised to include MCQ as assessment tools in summative examination for CBME batches from 2021 onwards as well.[5] With this background, the objectives of study were to analyze the quality of MCQs of examination by doing the Item analysis.

Aims and Objectives

The objectives are as follows:

1. To carry out item analysis of MCQs used in formative assessment to know the validity
2. To carry out a post validation item analysis of MCQ’s of 1st MBBS students in anatomy to use results for further actions.

**MATERIALS AND METHODS**

The first MBBS students were administered 45 MCQ with single best response type and four distractor for Anatomy subject 112 students appeared out of 125 students. MCQs were allotted one mark each for correct answer, no negative marking for the wrong answer the result of student’s performance was used to determine the level of difficulty and power of discrimination using Microsoft office Excel.

Post validation of the paper was done by item analysis. The scores of all the students were arranged in ascending order. The upper one-third students were considered high achievers and lower one-third as low achievers. Each item was analyzed for indices as.\(^4\)

Difficulty index (P) = \( h + \frac{l}{n} \times 100 \)

Discrimination index (D) = \( h - \frac{l}{n} \times 2 \).

Where;
- \( h \) = number of students answering correctly in high achievers group.
- \( l \) = number of students answering correctly in the low achievers group.
- \( n \) = total number of students in both groups including non-responders.

Interpretation of difficulty index (P):
- \( P \leq 30\% \) - difficult
- \( P = 30–70\% \) - Acceptable
- \( P \geq 70\% \) - easy.

Higher the difficulty index; the lower is the difficulty to avoid confusion facility index term can be used-lower the facility index-difficult is question, higher facility index suggest easy level of question.

Discrimination index (D):
- \( D \) = Negative suggests - Defective item/wrong key
- \( D = 0–0.19 \) - Poor discrimination
- \( D = 0.2–0.29 \) - Acceptable discrimination
- \( D = 0.3–0.39 \) - Good discrimination
- \( D \geq 0.4 \) - Excellent discrimination

For discrimination; higher the index suggest that item can discriminate high achievers and low achievers.\(^{6,7}\)

Nonfunctional distractor in an item is the option, other than the key selected by <5% of students and functional or effective distractor is the option selected by 5% or more students. This suggests the distractors are plausible and not dummy.\(^6\)

**Statistical Analysis**

The indices were calculated using formulae referred in methods. The data were reported as % and mean ± standard deviation of all items. The relationship between the item difficulty index and discrimination index values for all items was determined using Pearson correlation analysis. \( P < 0.05 \) was considered to indicate statistical significance.

**RESULTS**

In the present study, [Table 1] mean difficulty index (P) was 56.67 ± 22.09 and mean discrimination index (D) was 0.35 ± 0.23. According to difficulty index [Figure 1], 13% questions were difficult (\( P \leq 30\% \)), 67% were of moderate difficult (\( P = 30–70\% \)), and 20% were easy (\( P \geq 70\% \)). Discrimination index [Figure 2] suggested, 20% items were poor discriminators (D < 0–0.19), 20% were acceptable (D = 0.2–0.29), 27% good (D = 0.3–0.39), and 33% were excellent discriminators (D > 0.4) none of item was negatively discriminating. All of the distractor were functional. The discrimination index correlated positively with the difficulty index (\( r = 0.105, P = 0.64 \)) which is not significant statistically.

![Figure 1: Difficulty index and distribution](image1.png)

![Figure 2: Discrimination index and distribution](image2.png)
Teli and Kate Item analysis of multiple-choice questions

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Mean</th>
<th>SD</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difficulty index</td>
<td>56.67816753</td>
<td>22.09039735</td>
<td>17.6–90.5</td>
</tr>
<tr>
<td>Discrimination index</td>
<td>0.357142857</td>
<td>0.235420185</td>
<td>0.08–0.73</td>
</tr>
</tbody>
</table>

### DISCUSSION

Our study showed that 80% items had discrimination above acceptable level - 20% were acceptable (D = 0.2–0.29), 27% good (D = 0.3–0.39), and 33% were excellent discriminators while 20% items were poor discriminators (D < 0–0.19), (D > 0.4) none of item was negatively discriminating.

About 13% items were difficult (P ≤ 30%), 67% were of moderate difficult (P = 30–70%), and 20% were easy (P ≥ 70%) Too easy and too difficult items were found to be poor discriminators.

The results found in our study- too easy and too difficult items were found to be poor discriminators, are consistent with studies by Karelia et al., Afra Musa et al., Kolte et al., and Thorat et al. Pande et al., Shete et al. study by Karelia et al., Pande et al., Shete et al. showed difficulty index correlated positively with the discrimination index which was not significant statistically. In our study as well, the discrimination index correlated positively with the difficulty index (r = 0.105, P = 0.64) which is not significant statistically. Sim and Rasiah and Mitra studies showed poor correlation between difficulty index and discrimination index. We did not report any negatively discriminating items while Afra Musa et al. reported 0.8% out of total 645 items.

The differences in the correlation between the difficulty index and discrimination index, negatively discriminating items could be mainly attributed to variation in sample size. The number of MCQ items is small in our study 45 items as compared to study by Karelia et al. and Afra Musa et al. who have tested 200 items and 645 items, respectively. The difference in the MCQ construct, teaching learning experiences offered, learning preferences by the cohort and size of the cohort are the various aspects which can give rise to these differences.

A study by Tarrant on 2770 items for MCQ writing flaws suggested non adherence to framing guidelines, testing lower cognition were major issues.

**Item analysis helps us to decide which item finds place in question bank and which can be used after re-framing, reconstructing and evaluating. Questions which are too easy or too difficult function as less discriminating. Hence, these questions need to be reconstructed to moderate level of difficulty, either by changing stem or supplying better-plausible distractor which will not testing interpretative skill or language skill of students. Negative discrimination index (the poor students did better than the best students) means item is usually poorly constructed and the key provided may be wrong. Properly developed MCQ to suit a particular group of students will have moderate difficulty and high discrimination. Thus, difficulty and discrimination index serves as an indicator of functional quality of each item.**

**Strengths and Limitations**

Most of the items were of acceptable difficulty level and good to excellent discrimination level. No negatively discriminating item and absence of non-functional distractor suggest good quality framing of the MCQs. This is a good step towards building item banks.

Smaller number of items tested on one group of students is the main limitation of the study. More reliable results, teacher generated items testing higher cognitive skills of acceptable difficulty and good discrimination will be possible with continuous testing using more number of items, testing on multiple batches of students.

### CONCLUSION

In the study, most of items had moderate difficulty and good to excellent discrimination. Too easy and too difficult items showcased poor discrimination - these items could be modified before using for next examinations. Along with these findings, no negatively discriminating item and absence of non-functional distractor suggest good quality framing of the MCQs. Training of faculty not only in framing MCQ but also using results of item analysis is perceived as important aspect.

### REFERENCES

6. Ananthkrishnan N. The item analysis. In: Medical Education...

How to cite this article: Teli C, Kate N. Item analysis of multiple-choice questions in anatomy for first year MBBS. Natl J Physiol Pharmacol 2022;12 (Online First). DOI: 10.5455/njppp.2022.12.01029202227012022

Source of Support: Nil, Conflicts of Interest: None declared.