*RESEARCH AND CLINICAL MEDICINE JOURNAL*
*The European Journal of Innovative, Integrative and Translational Medicine*
*Volume VI, Issue 1, 2022*

# Analysis of Multiple-Choice Questions Based Medical Undergraduate Assessment Tool

Sheetal Diliprao Bhavsar [1], Sushma Shivaji Jadhav [1], Rahul Suryakantrao Abhange [2]

[1] Department of Physiology, [2] Department of Pathology
Vilasrao Deshmukh Government Medical College, Latur, Maharashtra, India.

## ORIGINAL PAPER

### Abstract

***Introduction.*** MCQs are increasingly being used to assess students' knowledge. However, MCQs should be with quality i.e. valid & reliable for efficient assessment. For the purpose of assessing the quality of items and developing valid question bank item analysis was done. ***Aim.*** By item analysis, the quality of MCQs is assessed using the difficulty index, discrimination index, and distractor efficiency. ***Material and methods***. A total 60 items with 180 distractors from an internal examination in Physiology, of 150 medical students of first phase of Vilasrao Deshmukh Government Medical college, Latur, were analysed by calculating the difficulty index(P), discrimination index (D) and distractor efficiency (DE). ***Results.*** The mean of P was 58.07±23.26%. Out of 60 items, 33 had acceptable & 06 had difficult P value. The mean of D was 0.17±0.13. Twenty-three had acceptable to excellent D. Thirty-four items were with poor D and 3 were having negative D indicating poor framing of items or under preparation of students. Out of 180 distractors, 56 (31.11%) were non-functioning distractors (NFDs). There were 26(43.33%) items with zero NFDs and only 3 items had zero FDs. Fifteen items had only one NFD (66.6% distractor efficiency) and 16 had two NFDs (33.3% distractor efficiency). Some items with NFDs found to have acceptable P value. ***Conclusion.*** The study concludes that item analysis is easy & effective way of analyzing items. So, after every examination item analysis is essential to improve standard of assessment to assess learner's knowledge.

**Keywords:** difficulty index, discrimination index, distractor efficiency

## I. INTRODUCTION

Assessment drives learning because it includes both formative and summative knowledge and has a strong influence on learning [1]. The medical education worldwide consists of an initial assessment of a learner's needs, monitoring of teaching-learning activities, certification of the competence to award a degree and practice medicine in context to the of the societal need[2].It gives insights about students' learning and competencies. It also provides active feedback to the teachers on their educational actions [3].

Multiple choice questions (MCQs) are the most used assessment tool for medical students [4]. Being with objectivity MCQs (items) avoid the assessor's bias during scoring, having the advantage of covering vast content and assessing many candidates [5]. But if not well constructed, it will direct students to guessing. It is twice as reliable compared to short-answer questions [6] & is found to be superior to the modified essay question in assessing higher-order skills [7].

An MCQ consists of a stem and possible options. The stem of an item can be in question form, or an incomplete statement, or in fill in the blanks form. An item mostly has four options, with one correct answer and three wrong options which act as distractors. The teacher needs to know how good the test questions are and whether the test items were able to reflect students' performance in the course related to learning as he constructs and assesses a test [8]. The constructed items should be valid & reliable, yielding the same results if administered to the same group of test takers multiple times [9]. The reason for poorly developed MCQ can be item writing flaws like ambiguous stems, negatively worded

***Corresponding author:*** Dr Rahul Suryakantrao Abhange , Associate Professor
Department of Pathology, Vilasrao Deshmukh Government Medical College Latur, Maharashtra, India.
E-mail: abhangerahul24@gmail.com

options, or options that are not plausible. All of them affect the quality of the MCQs and make the assessment process less reliable and valid [10].

Item analysis, also called post-validation analysis, is the statistical process of collecting, summarizing, and using information from students' responses to assess the quality of test items and the test as a whole [11]. The purpose of item analysis is to identify good MCQs, which is based on difficulty index (P) or *P*-value, discrimination index (D), and distractor efficiency (DE). It is a relatively simple but valuable procedure performed after the examination that provides information regarding the reliability & validity of a test [12]. The main aim of item analysis is to improve standard of tests by revising or eliminating ambiguous or misleading items.

The aim of the present study was to assess the quality of MCQs (items) by determining the difficulty index, discrimination index of items, & distractor efficiency.

## II. MATERIAL AND METHODS

This cross-sectional study was conducted in the department of Physiology, Vilasrao Deshmukh Government Medical College, Latur, Maharashtra in the month of January 2022 using term end & preliminary examination scores of 150 students of phase I MBBS of batch 2020–21. The study was approved by the institutional ethical committee.

The study analyzed 60 single response type MCQs. Each item has a single stem with four options, including one being a correct answer & three distractors (incorrect answers). Each item assigned one mark. The maximum possible score was 60, and the minimum possible score was zero, with no negative marking. Thirty-five percent was the passing score.

The data obtained was entered into MS Excel & analyzed. The scores of 150 students were arranged in descending order from highest to lowest marks. Then, students were divided into three groups. One group consisting of higher scores was considered as the higher score achievers (H) & another group consisting of lower scores as the low score achievers (L) group. Out of 150 students, 50 were in the H & 50 in the L groups, the rest 50 were in the middle group and were not included in the study. This serves the twin purpose of having groups large enough to be representative and different enough to be meaningful [13].

The students were ranked in order of merit using Microsoft Excel 2010. The bottom third (50) of low achievers (L) and the top third (50) of high achievers (H) were used for item analysis to serve the twin purposes of having groups large enough to be representative and different enough to be meaningful [14]. We analyzed a total of 60 items with 180 distractors. Based on this data, various indices such as the difficulty index (P), discrimination index (D), and distractor efficiency (DE) were calculated for all items using standard methodological formulae.

DIF I (P): It is determined by calculating the percentage of examinees that answered the item correctly.

$$\text{Difficulty Index (P)} = [(H+L)/N] \times 100$$

$$\text{Discrimination Index (D)} = [(H–L)/N] \times 2$$

Where: H – Number of students answering correctly in High score achievers' group

L – Number of students answering correctly in Low score achievers' group

N –Total Number of students in both groups

The difficulty index is an ease index because the higher the value of P, the easier an item is and vice versa [15].

DIF I (P):
Range: 0 – 100%
P< 30% : Difficult
P 30–70%: Acceptable
P 50–60%: Ideal
P >70% : Easy

A higher value of P shows that greater number of students have answered correctly. It indirectly proves that questions are easy to attempt [16].

The discrimination index (D) is the ability of an item to differentiate between high performing students and low performing ones. Negative D indicates defective items or wrong key.

Discrimination index (D):
Range: 0 – 1
≥0.36 : Excellent
0.25-0.35 : Good
0.21-0.25 : Acceptable
≤ 0.20 : Poor

DE is defined as the ability of incorrect answers to distract the students. A non-functional distractor (NFD) means < 5% students have chosen an incorrect answer & a functional distractor means an option that is selected by 5% or more students. Based on the number of NFDs in an item, DE would be 0%, 33.3%, 66.6% and 100% [17].

## III. RESULTS AND DISCUSSIONS

### *A. Results*

A total 60 MCQs & 180 distractors were analyzed. The score of 100 students ranged 24-47. The mean & standard deviations (SDs) of P (DIFI), D (DI**)** & DE were 58.07±23.26%, 0.17±0.13 & 68.86±31.8% respectively (Table 1).

Table 1: Characteristics of MCQs evaluation criteria [18]

| Parameters | Marks |
|---|---|
| Student (n) | 150 |
| MCQs (n) | 60 |
| Score Total (n) (mean±SD) | 60 |
| Score obtained (mean±SD)<br>Range | 35.67±5.78<br>24-47 |
| Difficulty index (%) (mean±SD)<br>Range | 58.07±23.26<br>2-98 |
| Discrimination index (mean±SD)<br>Range | 0.17±0.13<br>-0.3 - 0.46 |
| Distractor efficiency (%) (mean±SD)<br>Range | 68.86±31.8<br>0-100 |

Out of 60 items, 33 had an 'Acceptable' level of difficulty & 21 had an 'Easy' level of difficulty, while 6 were 'Difficult' items (Table 2, Fig.1).

| Table 2: Difficulty index (P)[19] | | |
|---|---|---|
| P | No. of items n=60 | Interpretation |
| <30% | 06(10%) | Difficult |
| 30-70% | 33(55%) | Acceptable |
| >70% | 21(35%) | Easy |

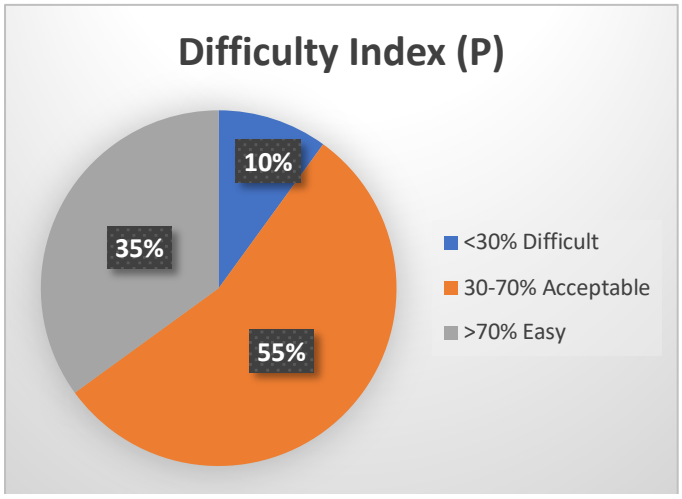


***Figure 1:*** *Proportion of difficulty index*

Out of 60 items, 11 had 'Good' & 8 had 'Acceptable' discrimination power. Four items had excellent discrimination power& 34 items had a poor discrimination index. Besides this, 3 items had negative discrimination index (Table 3, Fig. 2).

| Table 3: Discrimination Index (D) [19] | | |
|---|---|---|
| D | No. of items n=60 | Interpretation |
| ≥0.36 | 04 (6.67%) | Excellent |
| 0.25-0.35 | 11(18.33%) | Good |
| 0.21-0.24 | 08(13.32%) | Acceptable |
| ≤ 0.20 | 34(56.6%) | Poor |

Fifty-six out of 180 were non-functional distractors (NFDs). Twenty-six items were with zero NFDs means having 100% distractor efficiency (all functional distractors) while only 3 items had zero functional distractors (FDs), i.e. zero% distractor efficiency. Fifteen items were with only one NFD (66.6% distractor efficiency) & 16 had two NFDs i.e. with 33.3% distractor efficiency (Table 4).

Thirty-three items having an Acceptable level of difficulty index when analyzed for discrimination index 7 items had an Acceptable D but 14 were with oor D. Only 3 were Excellent while 9 were with Good D value (Table 5).

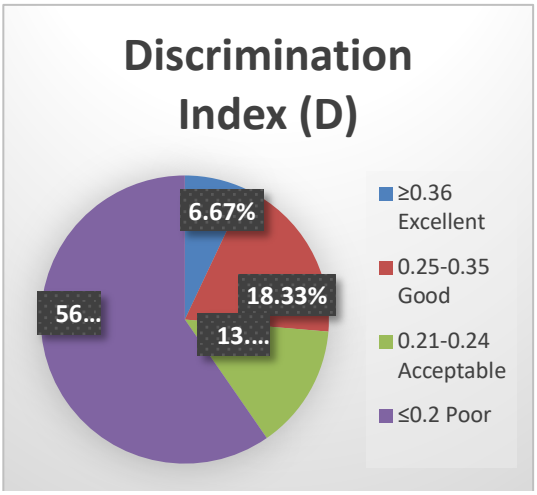


***Figure 2:*** *Proportion of discrimination index*

| Table 4: Analysis of Distractors [19] | |
|---|---|
| Parameters | Number/% |
| MCQs (Total) | 60 |
| Distractors (Total) | 180 |
| Functional Distractors | 124 (68.89%) |
| Non-Functional Distractors | 56 (31.11%) |
| MCQs with zero NFDs/ 4 FDs (DE=100%) | 26 (43.33%) |
| MCQs with 1 NFDs/ 3 FDs (DE=66.6%) | 15 (25%) |
| MCQs with 2 NFDs/ 2 FDs (DE=33.3%) | 16 (26.67%) |
| MCQs with 3 or more NFDs/1 or 0 FDs (DE=0%) | 03 (5%) |

| Table 5: Discrimination index of items with acceptable P value | |
|---|---|
| D | No. of items, n=33 |
| Excellent | 03(9.09%) |
| Good | 09(27.3%) |
| Acceptable | 07(21.2%) |
| Poor | 14(42%) |

There were 34 items with different numbers of NFDs, out of which 14 had an Acceptable difficulty index. Three items were with a less than 30% difficulty index while 17 were at the Easy level of P. And out of 34 items, 26 (76.47%) had a Poor discrimination index. Five (14.7%) items were with good discrimination index while 2 had Excellent discrimination index & only one had an Acceptable discrimination index (Table 6).

| Table 6: Difficulty index (P) & discrimination index (D) of items with different no. of NFDs[19] | | | |
|---|---|---|---|
| P | No. of items n=34 | D | No. of items n=34 |
| Difficult | 03(8.8%) | Excellent | 2(5.88%) |
| Acceptable | 14(41.2%) | Good | 5(14.7%) |
| Easy | 17(50%) | Acceptable | 1(2.94%) |
| -- | -- | Poor | 26(76.47%) |

*B. Discussion.*

The study was done to analyze items which are used to assess medical students' knowledge. Assessment is an essential part of learning process in education and MCQs are the most used tool. It quickly evaluates different levels of cognition like comprehension, application, analysis, and synthesis among students [20]. MCQs need to be of quality which means should be valid and reliable. Therefore, to assess items we did post-validation analysis using DIFI, DI& DE, which is a simple but effective method. It also detects specific technical flaws, provides feedback for further improvement [21] &helps to develop a pool of valid MCQs. An ideal MCQ should have average level of difficulty (>30- 60%) with a higher discrimination index (>0.25) and 100% distractor efficiency (means all three incorrect responses should function) [22].

The mean score obtained was 35.67±5.78 & 100% of students are having passing marks. Mean P was 58.07±23.26% which is in the acceptable range of difficult index(P).Out of 60 items 33 (55%) were well within the acceptable range of difficult index(P). Nine were ideal (P:50-60%). Thirty five percent (23) were easy& 10% (06) were in difficult range & should be revised or discarded. The difficult items should be reviewed for any confusing language, incorrect key or else discarded. The reason for items to be difficult may be because of had not been taught properly or difficult for students to grasp it or may not be studied it at all. These items can be used for selecting toppers. Easy questions too needed for increasing confidence level of students. Too difficult items (P ≤ 30%) will lead to deflated scores, while easy items (P>70%) will result into to inflated scores and a decline in motivation [23]. Only 35% of the total test items had difficulty index scores crossing 70%. This observation was quite close to a study of year two examinations of a medical school reported by Si Mui Sim et al (2006), who found that about 40% of the MCQ items crossed difficulty index 70% showing that the test items were easy for the examinees [24]. It is advisable to place easy items at the beginning of the paper to boost the confidence of overall students and taken as warm up questions. Difficult items can be used to select toppers [25].

Discrimination index differentiates between students with higher and lower capabilities. The theoretical range of D is from negative to 1.0, with >0.35 being acceptable: and higher the D more differentiating is an item [26].In our study mean D was 0.17±0.13. Eleven & 8 items had good and acceptable D respectively so these items can be retained &stored in the question bank. Four items were excellent to discriminate the students. An item may show low discrimination if the test covers a wide range of content areas at different taxonomic levels of cognitive skills (Mehrens& Lehmann, 1983) [27]. Thirty-four items had poor D which indicates that the items were too difficult or too easy and need to be revised or discarded. Three had negative D means poorly framed with incorrect key or generalized under preparation of students and should be discarded as these items reduces validity of items [28]. Some studies have shown negative D in 20% and 4% items [29,30]. Revising these items may contribute to the validity of the test on next test.

When items with acceptable difficulty level were analyzed for discrimination, 9 & 7 had good & acceptable discriminating power. Three were with excellent power to differentiate high score achievers from low score achievers. Though 14 were in acceptable difficulty level, having poor discriminating power, so these should be revised.

Analysis of the distractors identifies their errors, so that they may be revised, replaced or removed [31]. Distractors used in the items need to be plausible so that they attract at least 5% of students [32]. Hingorjo observed that items having one NFD had excellent discrimination ability [33].

Out of 180 distractors, 56 were non-functional. Items with 100% distractor efficiency that is with zero NFD, can be retained in the question bank. We should remove or replace NFDs with a more plausible option. More NFDs in an item increases P (makes an item easy) and reduces D. While framing quality MCQ, distractors should be plausible and the number of NFDs should be reduced. However, NFD is not the only factor that contributes to the difficulty of an item. Flaws in item writing also contribute to poor student performance [34]. Thirty-four items were with more than zero NFDs. We determined the difficulty level of each item having more NFDs before revision and found that 14 that is about 41.2% in the acceptable range of difficulty so can be retained. Therefore, items though have more NFDs, there is no need to discard them& can be retained. Three items having a difficult level of difficulty index need revision. Fifty percent were easy & should be discarded or revised.

## IV. CONCLUSION

Item analysis is an easy & effective method of assessing the quality of MCQs. It helps to identify items with good &poor indices. Depending upon analysis, we can decide whether to retain, revise or discard items & this way can improve the standard of MCQs. It is helpful to create a quality question bank. In this study, the majority items were in the acceptable range of difficulty index, so they can be stored in the question bank. But easy items also were there which need to be revised or discarded, otherwise it will create false confidence in students. Many items had poor discriminating power, hence it will be better to discard those items. Items with an acceptable difficulty level had poor discriminating power. Therefore, decision to retain or discard depends upon the purpose of the examination. Many items have NFDs which

need to be replaced or removed. Some MCQs, though having more NFDs have been found to have acceptable difficulty level so these items should be further analyzed for difficulty before revision. As most of the items with a greater number of NFDs had poor discriminating power their retention depends on the purpose of examination. The analysis of defective items should be done for flaws, and it should be optimized. These should not be blindly dropped; otherwise, some skills may be left out of the assessment.

The results show that there should be change in the way MCQs are being selected for any test. After every examination analysis need to be done to improve the standard of assessment as standardized MCQs only can efficiently assess learner's knowledge.

**Conflict of interest**: none to declare.

## REFERENCES


1. Saint DA, Horton D, Yool A, Elliott A. A progressive assessment strategy improves student learning and perceived course quality in undergraduate physiology. Adv Physiol Educ. 2015; 39(3):2()18-222
2. Ananthakrishnan N. Multiple choice questions facts and fantasies and Item Analysis validation and banking of MCQs. In: N Ananthakrishnan, K R Sethuraman, Santhosh Kumar, editors. Medical Education Principles and Practice. 2nd ed. Pondicherry: Alumni Association of National Teacher Training Centre JIPMER; 2000. pp. 119-37.
3. Cunnigham S. Can pediatric medical students devise a standard for their colleagues? Arch Dis Child 1999; 80: 573-575
4. Mehta G, Mokhasi V. Item Analysis of Multiple Choice Questions- An Assessment of the Assessment Tool. IJHSR. 2014;4(7):197-202.
5. McCoubrie P: Improving the fairness of multiple-choice questions: a literature review. Med Teach 2004, 26(8):709-712.
6. Kuechler WL, Simkin MG. How well do multiple choice tests evaluate student understanding in computer programming classes? J Inf Syst Educ. 2003;14(4):389-400.
7. Palmer EJ, Devitt PG. Assessment of higher-order cognitive skills in undergraduate education: Modified essay or multiple choice questions? Research paper. BMC Med Educ 2007;7:49
8. Skakun EN, Nanson EM, Taylor WC, Kling S. An investigation of three types of multiple choice questions. Ann Conf Res Med Educ 1977; 16: 111-116.
9. Verma S. 1999 Preliminary Item Statistics Using Point Biserial Correlation And P-Values. Education Data Systems Inc., Morgan Hills, USA
10. Islam ZU, Usmani A. Psychometric analysis of anatomy MCQs in modular examination. Pak J Med Sci. 2017;33 (5):1138-1143.
10. Collins, J, Writing Multiple-Choice Questions for Continuing Medical Education In Radiology, Activities and Self-Assessment Modules1 The Journal of continuing Medical Education, March ( 2006) 26, 543-551. PMID: 16549616
11. Scorepak®: Item analysis. [Last citied on 2013 Apr 13]. Available from:www.washington.edu./oea/score1/htm
12. Ananthakrishnan N. Multiple choice questions facts and fantasies and Item Analysis validation and banking of MCQs. In: N Ananthakrishnan, K R Sethuraman, Santhosh Kumar, editors. Medical Education Principles and Practice. 2nd ed. Pondicherry: Alumni Association of National Teacher Training Centre JIPMER; 2000. pp. 119-37.
13. Carneson J, Delpierre G, Masters K. Designing and managing MCQs. Appendix C: MCQs and Blooms taxonomy(Online) 2011(Cited 2011 Feb2). Available from URLhttp://web.uct.ac.za/projects/cbe/ mcqman/mcqappc. html.
14. Hingorjo MR, Jaleel F. Analysis of one-best MCQs: The difficulty index, discrimination index and distracter efficiency. J Pak Med Assoc 2012;62:142-7.
15. Chauhan P, Ratrhod S, Chauhan B, Chauhan G, Adhvaryu A, Chauhan A. Study of Difficulty Level and Discriminating Index of Stem Type Multiple Choice Questions of Anatomy in Rajkot, BIOMIRROR 2013; Volume 4:1-4.
16. Ananthakrishnan N. Item Analysis validation and banking of MCQs. In: Medical Education Principles and Practices. 2nd ed. Ananthkrishnan N, Sethuraman KR, kumar S. Alumini Association of National Teacher Training Centre, JIPMER, Pondichery, India. p,131-137.
17. Wajha M, Alam S, Hassan U et al.Difficulty Index, Discrimination Index and Distractor Efficiency in Multiple Choice Questions. Ann. Pak. Inst. Med. Sci. 2017;310-15
18. Patil R, Palve S, Vell K et al. Int J Community Med Public Health. 2016 Jun;3(6):1612-1616
19. Carneson J, Delpierre G, Masters K. Designing and managing MCQs. Appendix C: MCQs and Blooms taxonomy (Online) 2011. (Cited 2011 Feb2). Available from URLhttp. web uct ac za/projects/cbe/mcqman/mcqappchtm
20. Considine J, Botti M, Thomas S. Design, format, validity and reliability of multiple choice questions for use in nursing research and education. Collegian. 2005;12:19–24.
21. Patil R et al. Evaluation of multiple choice questions by item analysis in a medical college at Pondicherry, India .Int J Community Med Public Health. 2016 Jun;3(6):1612-1616
22. Eaves S, Erford B. The Gale group: The purpose of item analysis, item difficulty, discrimination index, characteristic curve. Available from: www.education.com/reference/article/itemanalysis/ [Last cited on 2013 Apr 15].
23. Si-Mui Sim, Rasiah RI. Relationship between item difficulty and discrimination indices in true/false type multiple choice questions of a para-clinical multidisciplinary paper. Ann Acad Med Singapore 2006; 35: 67-71.
24. Dr. Swapnagandha S. HalikarDr.VeerendraGodboleDr.Saurabh Chaudhari Item Analysis to Assess Quality of MCQs. Indian Journal Of Applied Research Volume : 6 | Issue : 3 | March 2016 | ISSN - 2249-555X | IF : 3.919 | IC Value : 74.50
25. Singh, T. &Anshu (2012) Principles of Assessment in Medical Education, New Delhi: Jaypee Brothers Medical Publishers.
26. 27.Mehrens, W & Lehmann, I (1983) Measurement and Evaluation in Education and Psychology, 3rd edn. New York: Holt, Rinehart & Winston.
27. Matlock-Hetzel S. Presented at annual meeting of the Southwest Educational Research Association, Austin, January; 1997. [Last cited on 2013 Apr 13]. Basic concept in item and test analysis. Available from:www. ericae.net/ft/tamu/espy.htm .
28. GajjarS, Sharma R, Kumar P,Rana M. Item and test analysis to identify quality Multiple Choice Questions (MCQs) from an assessment of medical students of Ahemdabad, Gujarat. Indian journal of Community Medicine 2014;39:17-2
29. Hingorjo MR, Jaleel F. Analysis of one-best MCQs: The difficulty index, discrimination index and distracter efficiency. J Pak Med Assoc. 2012;62:142–7. [PubMed]
30. Dufresne RJ, Leonard WJ, Gerace WJ. Making sense of students answers to multiple choice questions. Phys Teach 2002;40:174-180
31. Parag Chavda et al. Item Analysis of Multiple Choice Questions based Undergraduate Assessment in Community Medicine. South East Asian Journal of Medical Education Vol. 9 no. 1, 2015:66-68
32. Hingorjo MR, Jaleel F. Analysis of one-best MCQs: The difficulty index, discrimination index and distractor efficiency. J Pak Med Assoc 2012;62:142-7.
33. Abdulghani HM, Ahmad F, Ponnamperuma GG, Khalil MS, Aldrees A. The relationship between non-functioning distractors and item difficulty of multiple choice questions: A descriptive analysis. J Health Spec. 2014;2:148-51.