International Journal of Research in Medical Sciences
Garg R et al. Int J Res Med Sci. 2019 Jan;7(1):174-177
www.msjonline.org
pISSN 2320-6071 | eISSN 2320-6012
DOI: http://dx.doi.org/10.18203/2320-6012.ijrms20185375

Original Research Article

Analysis of multiple choice questions from a formative assessment of medical students of a medical college in Delhi, India

Richa Garg[1], Vikas Kumar[2]*, Jyoti Maria[1]

[1]Department of Pharmacology, [2]Department of Community Medicine, Dr. Baba Saheb Ambedkar Medical College & Hospital, Delhi, India

Received: 17 October 2018
Accepted: 13 November 2018

***Correspondence:**
Dr. Vikas Kumar,
E-mail: drvikaspsm@gmail.com

ABSTRACT

Background: Assessment is a dominant motivator to direct and drive students learning. Different methods of assessment are used to assess medical knowledge in undergraduate medical education. Multiple choice questions (MCQs) are being used increasingly due to their higher reliability, validity, and ease of scoring. Item analysis enables identifying good MCQs based on difficulty index (DIF I), discrimination index (DI), and distracter efficiency (DE).
Methods: Students of second year MBBS appeared in a formative assessment test, that was comprised of 50 "One best response type" MCQs of 50 marks without negative marking. All MCQs were having single stem with four options including, one being correct answer and other three incorrect alternatives (distracter). Three question paper sets were prepared by disorganizing sequence of questions. One of the three paper sets was given to each student to avoid copying from neighboring students. Total 50 MCQs and 150 distracters were analyzed and indices like DIF I, DI, and DE were calculated.
Results: Total Score of 87 students ranged from 17 to 48 (out of total 50). Mean for difficulty index (DIF I) (%) was 71.6±19.4. 28% MCQs were average and "recommended" (DIF I 30-70%). Mean for discrimination index (DI) was 0.3±0.17. 16% MCQs were "good" and 50% MCQs were in "excellent" criteria, while rests of the MCQs were "discard/poor" according to DI criteria. Mean for distracter efficiency (DE) (%) was 63.4±33.3. 90% of the items were having DE from 100 to 33%. It was found that MCQs with lower difficulty index (≤70) were having higher distracter efficiency (93.8% vs. 6.2%, p=0.004).
Conclusions: Item analysis provided necessary data for improvement in question formulation and helped in revising and improving the quality of items and test also. Questions having lower difficulty index (≤70) were significantly associated with higher discrimination index (≥0.15) and higher distractor efficiency.

Keywords: Assessment, Difficulty index, Discrimination index, Distractor efficiency

INTRODUCTION

Assessment is an essential part of the learning process in medical education. For students, assessment is a dominant motivator to direct and drive their learning. Different methods of assessment namely Multiple Choice Questions (MCQs), Short Essay Questions (SEQs), Objective Structured Practical Examination (OSPE), Objective structured clinical examination (OSCE) and VIVA VOCE are commonly used to assess medical knowledge in undergraduate medical education.[1] Multiple choice questions (MCQs) or "items" are being used increasingly due to their higher reliability, validity, and ease of scoring.[2,3]

Formative assessments additionally give opportunity to the teachers to improve the weak areas of students.[4] Proper analysis of assessment allows the teachers to

conduct it accurately.[1] "Item analysis" examines student responses to individual test items (MCQs) to assess the quality of those items and test as a whole to improve/revise items and the test.[5-7] A good item can assess cognitive, affective, as well as psychomotor domain and is preferred over other methods. MCQ based evaluation apart from assessing knowledge also evaluates understanding and analyzing power of students.[4] Item analysis enables identifying good MCQs based on difficulty index (DIF I), discrimination index (DI), and distracter efficiency (DE).[5,7-11]

Difficulty index (DIF I) is denoted by FV (facility value) or P value.[1] DIF I describe the percentage of students who answered the item correctly and ranges between 0 and 100%.[1,3,4,6] DIF I is a misnomer as bigger is the value of DIF I, easier is the item and vice versa; hence, it is also called by some authors as ease index.[12] The higher percentage of score reflects that item is easier for students.

Whereas, Discrimination index (DI) describes the ability of an item to distinguish between high and low scoring students. It ranges between 0 and 1. The higher score reflects the excellent ability of item to discriminate between high and low performing students. DI of 1 is ideal as it refers to an item which perfectly discriminates between students of lower and higher abilities.[12] There are instances when the value of DI can be <0 (negative DI) which simply means that the students of lower ability answer more correctly than those with higher ability. Such situations though undesirable, happen due to complex nature of item making it possible for students of lower ability to select correct response without any real understanding. Here a student of lower ability by guess select correct response; while a good student suspicious of an easy question, takes harder path to solve and end up to be less successful.[7]

Distracter efficiency is one such tool that tells whether the distracters in item (MCQ) was well constructed or failed to perform its purpose in distracting students from selecting correct answer. Any distracter that has been selected by less than 5% of the students is considered to be a non-functioning distracter (NFD).[12]

The present study was conducted with the objective to assess the quality of MCQs (Item) with valid tools like difficulty index (DIF I), Discrimination Index (DI), and Distracter Efficiency (DE) and improve the MCQs to create a question bank for further assessment.

METHODS

The proposed study was an observational, cross sectional study. Total 87 out of 98 students of 2nd year MBBS, appeared in a formative assessment test after completion of unit "Autacoids" in Pharmacology. Assessment test was comprised of 50 "One best response type" MCQs of 50 marks. All MCQs were having single stem with four options including, one being correct answer and other three incorrect alternatives (distracter). Three question paper sets were prepared by disorganizing sequence of questions. One of the three paper sets was given to each student to avoid copying from neighbouring students.

Time allowed was one hour and each correct response was awarded 1 mark, no negative marking for incorrect response. All MCQ answer sheets were collected from students and data obtained were entered in MS Excel 2007. Each MCQ was analyzed with three tools that is Difficulty Index (DIF I), Discrimination Index (DI) and Distracter Efficiency (DE). Scores of students were entered in descending order and whole group was divided in three groups, upper 1/3 (higher ability group- HAG), middle 1/3 and lower 1/3 (lower ability group -LAG). As per the need for calculating the indexes, data related to higher and lower ability group were used in analysis.[12,6,13] Total 50 MCQs and 150 distracters were analyzed and indices like DIF I, DI, and DE were calculated with following formulas:

Difficulty Index (DIF I) = [(H + L)/N] × 100

Value of DIF I: 0 to 100%; where <30% = too difficult, 30%-70%= recommended, >70%= too easy[6]

Discrimination Index (DI) = 2 × [(H-L)/N]

Value of DI: 0 to 1; <0.15= Poor/ Discard, 0.15 to <0.25= Good, ≥0.25= Excellent [4]

Distracter Efficiency (DE) = M/N x 100

Value of DE: 0 to 100%; where <5% = poor (NFD)[13]

Where,

N: Total number of students in both upper 1/3 and lower 1/3 groups,
H: Number of students answering the item correctly in higher ability group (HAG),
L: Number of students answering the item correctly in lower ability group (LAG),
M: Number of students from both groups who choose that particular distracter.

RESULTS

Total 50 MCQs were analyzed with three different indices that is Difficulty Index (DIF I), Discrimination Index (DI) and Distracter Efficiency (DE). Score of 87 students ranged from 17 to 48 (out of total 50).

Difficulty Index (DIF I): Mean and standard deviation for DIF I (%) was 71.6±19.4; 14 out of 50 MCQs were average and "recommended" (30-70%) and rests of the MCQs were "too easy" or "too difficult" according to Difficulty Index (Table 1).

Table 1: Distribution of items according to difficulty index (DIF I) and discrimination index (DI).

Cut off points	Items (N=50)	Interpretation	Action
Difficulty index (P)			
<30	2	Too difficult	
30- 70	14	Age	Recommended
>70	34	Too easy	
Discrimination index (DI)			
<0.15	17	Poor	Discard/ revise
0.15-<0.25	8	Good	Store
≥0.25	25	Excellent	Store

Discrimination Index (DI): Mean and standard deviation for DI was 0.3±0.17. On analyzing all the MCQs by DI, 8 out of 50 MCQs were "good" and 25 MCQs in "excellent" criteria, while rests of the MCQs were "discard/poor" according to DI criteria (Table 1).

Distracter Efficiency (DE): Mean and standard deviation for DE (%) was 63.4±33.3; there were 55 non-functional distracters out of 150 distracters in total 50 MCQs. 90% of the items were having DE from 100 to 33% (Table 2).

Table 2: Distribution of items according to distractor efficiency.

Distractor analysis	Number	%
Number of items	50	
Number of distractors	150	
Non functional distractors (NFD)	55	36.7
Functional distractors (FD)	95	63.3
Items with 0 NFD (DE=100%)	17	34
Items with 1 NFD (DE=66.6%)	16	32
Items with 2 NFD (DE=33.3%)	12	24
Items with 3 NFD (DE= 0%)	5	10

To see the associations among Difficulty Index, Discrimination Index and Distracter Efficiency, Fisher's Exact and Chi-Square tests were applied. It was found that MCQs with lower Difficulty Index (≤70) were having higher Distracter Efficiency (93.8% vs. 6.2%, p=0.004). MCQs with higher Discrimination Index (≥0.15) were having higher Distracter Efficiency (81.8% vs. 18.2%, p=0.001). MCQs with lower Difficulty Index (≤70) were having higher Discrimination Index (81.2% vs. 18.8%) but association was found statistically insignificant (p=0.2) (Table 3).

Table 3: Association among difficulty index, discrimination index and distracter efficiency.

Indices		Distracter Efficiency		Total	Tests of Significance
		≥66% DE	**≤33% DE**		
Difficulty Index	≤70	15 (93.80%)	1 (6.20%)	16 (32.0%)	Fisher's Exact Test, p=0.004
	>70	18 (52.90%)	16 (47.10%)	34 (68.0%)	
Discrimination Index	<0.15	6 (35.30%)	11 (64.70%)	17 (34.0%)	Chi-Square= 10.82, Df=1, p=0.001
	≥0.15	27 (81.80%)	6 (18.20%)	33 (66.0%)	
		Discrimination Index			Fisher's Exact Test, p=0.2
		≥0.15	<0.15		
Difficulty Index	≤70	13 (81.20%)	3 (18.80%)	16 (32.0%)	
	>70	20 (58.80%)	14 (41.20%)	34 (68.0%)	
	Total	33 (66.0%)	17 (34.0%)	50 (100.0%)	

DISCUSSION

The present item analysis uses single best response type MCQ questions. In this study mean and standard deviation of difficulty index {DIF I (%)} was 71.6±19.4. Fourteen (28%) out of fifty MCQs were average and "recommended" (30-70%), 34 (68%) were too easy (>70%), 2 (4%) were too difficult (<30). Mean Discrimination index (DI) was 0.3±0.17, 8 (16%) items were in acceptable range, 25(50%) items in recommended or excellent criteria and 17 (34%) were poor. Mean for Distractor efficiency (DE) was 63.4±33.3, 95 (63.33%) were functional distracters and 55 (36.7%) were non functioning distractors (NFD) out of 150.

In a study conducted Patel R on 83 MBBS students of microbiology for 40 MCQs, reported mean DIF I of 55.9±15.7%, 12 (30%) items were in the ideal (50-60%), 18(45%) items in acceptable range (30-70%), 7(17.5%) items were easy (>70%) and 3 (7.5%) items were difficult (<30%).[14] Mean Discrimination index was 0.29±0.20, 17 (42.5%) items were excellent, 7 (17.5%) items were good, 1 (2.5%) item was acceptable and 15 (37.5%) items were poor. Mean for Distractor efficiency (DE) reported was 84.94±22.58%, Out of 120 distractors analysed, 102 (85%) were functional and 18 (15%) were non-functional.

In another study conducted by Saxena S et al, on 80 MBBS students of Biochemistry for 30 MCQs, mean DIF I of 43.42±18.68 was reported.[15] 21 out of 30 MCQs were in acceptable range, 1 was too easy and 8 were too difficult. Mean Discrimination index (DI) of 0.21±0.11 was reported, 8 (26.67%) items were in recommended DI

value, 16 (53.33%) items in acceptable range, and 6 (20%) items were reported poor. Mean DE of 95.55±11.55was reported, 26 (86.67%) items were with all functional distracters, and 4 (13.33%) items were with single non-functional distractor. Kolte, reported mean DIF I as 57.92±19.58.[16] In this study, the *P* value of 26 (65%) items was in acceptable range (30-70%), 10 (25%) items were easy (P >70%), and 4 (10%) items were difficult (P <30%). DI of 60% items was excellent (d value >0.35). 47.5% items had100% Distracter Efficiency (DE) and 7.5% items had 0% DE.

Item analysis done by Mehta et al, on 100 MBBS students for 50 MCQs of anatomy reported mean DIF I of 63.06±18.95 with 31 (62%) items in the acceptable range, 16 (32%) items were too easy and 3 (6%) items were too difficult.[17] Mean DI reported was 0.33±0.18. Out 50 items, 15 (30%) items had DI <0.2 (poor), 9 (18%) items was DI ≥0.20 and ≤0.35(Good), and 26 (52%) items had DI >0.35 (excellent). Out of total 150 distracters, 53 were NFD, 28 were functional distracters, and 69 had none response with mean DE of 63.97±33.56. In study by Gajjar et al, on 50 items with 150 distractors, 24 had "good to excellent" DIF I (31-60%) and 15 had "good to excellent" DI (>0.25).[4] Mean DE was 88.6% considered as ideal/ acceptable and Non Functional Distractors (NFD) were only 11.4%. Mean DI was 0.14. 133 were functional distracters and 17 were NFDs with mean DE of 88.6±18.6.

CONCLUSION

The items having average difficulty and good discriminating index with functional distractors should be used in further assessment tests for good quality evaluation. Item analysis provided necessary data for improvement in question formulation and helped in revising MCQs with poor discrimination index and thus improved the quality of items. Association of Difficulty index with distractor efficiency and discrimination index was significant. Item analysis should be incorporated into the process of test development and review. Estimation of indices along with finding association between them is recommended to develop the best possible question bank for assessment of medical students.

Funding: No funding sources
Conflict of interest: None declared
Ethical approval: Not required

REFERENCES

1. Mehta M, Banode S, Adwal S. Analysis of multiple choice questions (MCQ): important part of assessment of medical students. Int J Med Res Rev. 2016;4(2).
2. Case S, Swanson D. Constructing written test questions for the basic and clinical sciences. 3rd ed. Philadelphia: National Board of Medical Examiners; 2003.
3. Tarrant M, Ware J. A framework for improving the quality of multiple-choice assessments. Nurse Educ. 2012;37(3):98-104.
4. Gajjar S, Sharma R, Kumar P, Rana M. Item and test analysis to identify quality multiple choice questions (MCQs) from an assessment of medical students of Ahmedabad, Gujarat. Indian journal of community medicine: official publication of Indian Association of Preventive and Social Medicine. 2014 Jan;39(1):17.
5. University of Washington. Office of Educational Assessment. Understanding Item Analyses. Available at: http://www.washington.edu/assessment/scanning-scoring/scoring/reports/item-analysis/. Accessed 20 December 2018.
6. Singh T, Gupta P, Singh D. Principles of Medical Education. 3rd ed. New Delhi: Jaypee Brothers Medical Publishers (P) Ltd. Test and item analysis; 2009:70.
7. Matlock-Hetzel S. Basic concept in item and test analysis. Presented at annual meeting of the Southwest Educational Research Association, Austin, 1997. Available at: www.ericae.net/ft/tamu/Espy.htm.
8. Eaves S, Erford B. The Gale group. The purpose of item analysis, item difficulty, discrimination index, characteristic curve. Available at: www.education.com/reference/article/itemanalysis/
9. Sarin YK, Khurana M, Natu MV, Thomas AG, Singh T. Item analysis of published MCQs. Indian Pediatr. 1998;35:1103-5.
10. Tarrant M, Ware J, Mohammed AM. An assessment of functioning and non-functioning distractors in multiple-choice questions: A descriptive analysis. BMC Med Educ. 2009;9:1-8.
11. Scantron Guides-Item Analysis, adapted from Michigan State University website and Barbara gross devils tools for teaching. Available at: www.freepdfdb.com/pdf/item-analysis-scantron.
12. Hingorjo MR, Jaleel F. Analysis of one-best MCQs: the difficulty index, discrimination index and distractor efficiency. J Pak Med Assoc. 2012 Feb;62(2):142-7.
13. Cizek GJ, O'Day DM: Further investigations of nonfunctioning options in multiple choice test items. Educ psycho meas. 1994;54(4):861-72.
14. Patel R. Use of item analysis to improve quality of multiple choice questions in II MBBS. J Edu Technol Heal Sci. 2017;4(1):22-9.
15. Saxena S, Srivastava P, Mallick A, Joshi H, Das B. Item analysis: An unmatched tool for validating MCQs in Medical Education. Ind J Basic Appl Med Res. 2016;5(4):263-9.
16. Kolte V. Item analysis of multiple choice questions in physiology examination. Indian J Basic Appl Med Res. 2015;4(4):320-6.
17. Mehta G, Mokhasi V. Item analysis of multiple choice questions-an assessment of the assessment tool. Int J Health Sci Res. 2014;4(7):197-202.

Cite this article as: Garg R, Kumar V, Maria J. Analysis of multiple choice questions from a formative assessment of medical students of a medical college in Delhi, India. Int J Res Med Sci 2019;7:174-7.