

A Study on Quality check on Multiple Choice Questions (MCQs) Using Item Analysis for Differentiating Good and Poor Performing Students

Jatin Chhaya¹, Hitesh Bhabhor², Jayna Devalia³, Ullas Machhar⁴, Abhay Kavishvar⁵

¹Assistant professor, Community Medicine Department, Smt B. K. Shah Medical College, Sumandeep Vidyapeeth, Waghodia, Gujarat, India

²Tutor, ⁵Associate Professor, Community Medicine Department, Government Medical College, Surat, Gujarat, India

³Tutor, Physiology Department, Smt B. K. Shah Medical College, Sumandeep Vidyapeeth, Waghodia, Gujarat, India

⁴Assistant Professor, Microbiology Department, Stem cell Therapy Institute, New civil hospital, Surat, Gujarat, India

Correspondence : Dr. Hitesh Bhabhor, E mail: hitesh1357@yahoo.co.in

Abstract :

Introduction : Single best answer MCQs (Multiple Choice Questions) are becoming a popular form of formative as well as summative assessment in educational institutes, not only but including medical field. Formation of MCQs with appropriate level of difficulty and excellent level of discrimination power is a huge task and failing in it, results in failure of test and by that we end up with poor discrimination among high and low performing students. Difficulty index (p-value) and Discrimination Indices (DI) are tools that enable us to choose appropriate MCQs while planning the assessment of students. **Method :** The cross-sectional study was conducted in the department of Community Medicine of Government Medical College, Surat in June-July 2016, with 103 medical students of third year. As a routine practice formative assessment was planned with 40 MCQs, each have 4 options with one single best response as a key answer. The MCQs were analyzed for difficulty index, discrimination index. **Result :** Analyzed item with p-value and DI showed half of the items were at appropriate difficulty level (p-value between 30% - 70%), with excellent discrimination power (>0.24) and substantial proportion of item had poor discrimination ability. Two items found to have negative discrimination value while another two had zero discrimination power. **Conclusion :** Item analysis is an excellent tool that help us in formation of proper MCQs that have average Difficulty Index (DIF) with excellent Discrimination Index (DI). Poorly formed MCQs should be revised or discarded for future assessment.

Key words: Difficulty index, Discrimination index, Item analysis, MCQs

Introduction :

“Multiple Choice Questions (MCQs)” is a common practice to assess the knowledge of the MBBS students as a class room exam and even as an entrance test for post graduation. It made up of a question and a set of options. Each question has only one correct answer and usually, right answer gets set number of marks towards total marks while wrong answer earns nothing or it will be penalized, to discourage guessing.^[1,2]

MCQs are superior over the written exam as, (1) it can be scored rapidly and (2) it provides quick result (3) it is proved efficient to assess the knowledge of

large number of students together and (4) it is also possible to test the candidates at the desired level of difficulty by varying the difficulty of questions.^[3] Written test has lots of drawback like; (1) handwriting of students is not neat and clean (2) way of presentation of students has lots of differences (3) supplementary use is very high as compared to MCQs to evaluate status of children's knowledge and (4) teachers' appreciation over written answer is quite different.^[4] Bloom's taxonomy also mention that well framed MCQs are able to access cognitive domains such as analysis, synthesis and application of knowledge, instead of just remembering isolated information.^[5]

The purpose of MCQs is to differentiate students into good and poor performing students. So it becomes utmost important that quality of MCQs is also maintained at high standard, because poorly framed MCQs may be confusing evaluation tools, and it affects performance of high ability students.^[6] So, before giving MCQs for formative or summative assessment it is good to find out that; (1) How much are you sure that the MCQs are appropriate level of difficulty? (2) How will you know that the test efficiently separates between students who do well on the overall test and those who do not?

The long term objective of this study was to frame good quality of MCQs to set up a question bank for reference in future.

Method:

As a routine practice, formative assessment of 103 students of 3rd year part I MBBS was held in the subject of Community Medicine at the Department of Community Medicine, Government Medical College, Surat, Gujarat. There were 40 MCQs, each have 4 options with one single best response as a key answer. One mark was awarded for each correct answer and there was no penalty (negative markings) for incorrect response. The time allotted was one minute for each MCQ. Students were instructed that any discussion or otherwise inappropriate communication between them was not accepted. After the completion of exam, answer sheets were checked and data was entered in to excel sheet. Score of the students set in highest to lowest order. Afterward, the 27% of the students at the top and the 27% at the bottom were separated for the analysis, which was called high and low achieving students respectively.

Item Analysis was used to analyze:^[7-9]

- **Difficulty index (DIF or p-value):** It was also called ease index, and used to determine whether the item was of suitable level of difficulty for the batch of students tested. It ranges between 0 and 100%. Difficulty index was a misnomer as bigger was the value, easier was the item and vice versa.

$$p \text{ value} = [(H + L)/N] \times 100$$

Where, H - Correct response given by high performing students

L - Correct response given by low performing students

N - Total number of students in both groups

Interpretation of p-value as under:

- p-values > 70% - very simple questions.
- p-values < 30% - very hard items.
- p-value is between 30% to 70% - average items.
- **Discrimination index (DI):** It described the capability of a question to discriminate between students of good and poor performing. It ranged between 0 and 1. Higher the value of DI, question was more proficient to discriminate between students of higher and lower abilities.

$$DI = 2 \times [(H - L)/N]$$

Where, H - Correct response given by high performing students

L - Correct response given by low performing students

N - Total number of students in both groups

Interpretation of discrimination index was as under:

- 0.35 or higher - Excellent discrimination ability
- 0.25 to 0.34 - Good discrimination ability
- 0.15 to 0.24 - Marginal discrimination ability
- < 0.15 - poor items

Ethical consideration: ethical permission for the study was not taken from the institute. Our study was insight to change assessment of students and no any harmful technique was required to complete entire study.

Results:

A total of 103 students appeared in the formative assessment which consisted of 40 - single best

Table 1 : Descriptive analysis of 40 one-best answer MCQs

Sr no	Item description	Values
1	Total number of students	103
2	Total MCQs	40
3	Minimum score obtained for correct answer	10
4	Maximum score obtained for correct answer	33
5	Total mean score achieved	23 ± 4.52
6	Mean score of higher achieving students	28.5 ± 1.91
7	Mean score of low achieving students	17.57 ± 2.42
8	Mean of difficulty Index (DIF)	57.62 ± 25.14
9	Mean of discrimination Index (DI)	0.27 ± 0.22

Table 2 : Difficulty index and discrimination index of the 40 one-best respond MCQs

Q. No.	p-value	DI		Q. No.	p-value	DI
1	34	0.39		21	52	0.54
2	36	0.00		22	88	0.04
3	61	0.50		23	39	0.00
4	29	0.21		24	34	0.61
5	71	0.36		25	82	0.36
6	96	0.07		26	93	0.07
7	41	0.6		27	57	0.25
8	39	0.14		28	63	0.46
9	14	0.21		29	41	0.61
10	11	-0.07		30	80	0.32
11	91	0.11		31	50	0.50
12	57	0.57		32	84	0.11
13	41	0.39		33	91	0.11
14	63	0.32		34	16	0.04
15	61	0.79		35	59	0.39
16	64	0.21		36	91	0.11
17	75	0.36		37	91	0.11
18	16	0.25		38	59	0.46
19	18	-0.29		39	88	0.11
20	63	0.39		40	66	0.25

Table 3 : Assessment of 40 MCQs based on Difficulty index

Difficulty index	No.	%	Action
> 70% (too easy)	13	32.5	Revised/ Discard
30% - 70% (average difficulty)	21	52.5	Store
< 30% (too hard)	6	15	Revised/ Discard
Total	40	100	-

Table 4 : Assessment of 40 MCQs based on Discrimination index

Discrimination Index	No	%	Action
≥ 0.35 (Excellent discrimination)	17	42.5	Store
0.25 - 0.34 (Good discrimination)	5	12.5	Store
0.15 - 0.24 (Marginal discrimination)	3	7.5	Revised/ Discard
<0.15 (Poor discrimination)	15	37.5	Revised/ Discard
Total	40	100	

Table 5 : Inter-relationship of difficulty index and discrimination index for 40 one-best respond MCQs

Discrimination Index	Difficulty Index		
	too hard	average	too easy
$D < 0.00$	2 (revise/ discard)	2 (revise/ discard)	0 (revise/ discard)
$0 < D < 35$	4 (revise/ discard)	5 (store)	10 (revise/ discard)
$D \geq 35$	0	14 (store)	3 (store)

response from the subject of community medicine. Descriptive analysis for each item consist of range of correctness, mean score in upper, middle and lower group, mean of DIF and DI given in table 1.

(Table 2) The p-value and DI were calculated for each MCQ. Question no. 10 and 19 has negative discrimination value while question no. 2 and 23 has zero or no discrimination ability.

(Table 3) The difficulty index was worked out and it showed that 52.5% of question had moderate or average difficulty index (30% - 70%). It also revealed that 15% of MCQs were too much hard while 32.5% MCQs were considered too easy.

(Table 4) Out of 40 MCQs, 22 had excellent to good

discrimination ability, while 18 had marginal to poor discrimination ability.

(Table 5) The inter-relationship of the two index revealed that 22 (55%) items were 'perfect' having a p-value from 30 to 70, as well as a DI > 0.24.

Discussion:

One correct response out of many type of MCQ is most efficient tool for formative assessment of medical students; however this competence exclusively depends on quality of MCQ. Properly constructed multiple choice questions assess higher-order cognitive processing of Bloom's taxonomy such as interpretation, synthesis and application of knowledge, instead of just testing recall of isolated

facts.^[5] Poorly framed MCQs were not able to distinguish students who performed well from their counterpart and our test became a failure attempt. Item analysis is one of those tools that would protect us from that failure. Tarrent and Ware also validated that flawed MCQ items affected the performance of high-achieving students more than borderline students.^[10]

In the present study, the mean p value was 57.62 ± 25.10 , which was in the range of excellent level of difficulty (p=30 to 70 %). Our finding was corroborative with that of Mukherjee P and Mehta G who reported mean p value of 61.92 and 63.06 respectively.^[11,12] However few studies have reported lower mean p values.^[13,14]

DI is ability of item to distinguish between students of higher and lower abilities. Noteworthy, more difficult questions were answered incorrect by every student and too easy question answered correct by each student. So this type of questions has no discrimination power and we have to remove such type of questions from the test. Add to that, the question with negative discrimination index produce inverse result that it can be answer right by poor performing student and wrong by good performing student. Wrong answer key, vague framing of questions or generalized poor preparations of students are most responsible explanation for negative discrimination index. Further, item of having zero DI that means either it answered right by all or answered wrong by all, needs to be removed from assessment.

The mean DI found in this study was 0.27 ± 0.20 which is considered practically well, yet considerable proportion of items (47.5%) had poor DI. Earlier studies of Mukherjee P and Hingorjo MR reported mean of DI 0.31 ± 0.27 and 0.46 ± 0.08 respectively.^[11,15]

One study in our literature reported lower mean DI of 0.14 ± 0.19 .^[8] Items with DI > 0.35 were 42.5%, DI between 0.25 and 0.34 were 10%, DI 0.15 - 0.24 were 10% and DI <0.15 were 37.5%, which shown that almost 47.5% of MCQ in our study need to be

revised. Earlier study of Hingorjo MR^[15] reported, items with DI > 0.35 were 62%, DI between 0.25 and 0.34 were 14%, DI 0.15 - 0.24 were 12% and DI <0.15 were 12% (24% of MCQ need to be revised). Two items (4%) in our study had negative DI and two items (4%) had zero DI. Few studies reported higher proportion of negative DI.^[11,13,15]

We found 22 items (44 %) to be 'ideal' having a good p value (30 to 70%), as well as good to excellent DI (≥ 0.25). Hingorjo MR^[15] found 32 items (64%) as ideal having a p - value from 30 to 70, as well as a DI > 0.24. Other researchers have reported 24%, 30% and 46% items to be ideal in their studies.^[11-13]

Conclusion :

Items analyzed in the study showed that half of the items in assessment tool had poor or marginal discrimination power and half items had average difficulty level with excellent discrimination index. Results from this study highlighted the importance of item analysis. Items having average difficulty and excellent discrimination should be incorporated into forthcoming tests. This would also improve the general test score and appropriately discriminate among the students

Declaration:

Funding: Nil

Conflict of Interest: Nil

References :

1. Cizek GJ, O'Day DM. Further investigations of nonfunctioning options in multiple-choice test items. *Educ Psychol Meas* 1994; 54(4):861-72.
2. Skakun EN, Nanson EM, Kling S, Taylor WC. A preliminary investigation of three types of multiple choice questions. *Med Educ* 1979;13:91-6.A
3. Khan M-Z, Aljarallah BM. Evaluation of Modified Essay Questions (MEQ) and Multiple Choice Questions (MCQ) as a tool for Assessing the Cognitive Skills of Undergraduate Medical Students. *International Journal of Health Sciences*. 2011;5(1):39-43.
4. McAllister, D., and Guidice, R.M. (2012). Advantages and Disadvantages of Different Types of Test Questions. *Teaching in Higher Education*, 17 (2),193-207. [internet] Available from: <https://www.facultyfocus.com/articles/educational-assessment/advantages-and-disadvantages-of-different-types->

- of-test-questions/
5. Carneson J, Delpierre G, Masters K. Designing and managing MCQs: Appendix C: MCQs and Bloom's taxonomy. (Online) 2011 (Cited 2017 Sep 2). Available from URL: <http://web.uct.ac.za/projects/cbe/mcqman/mcqappc.html>.
 6. Christian DS, Prajapati AC, Rana BM, Dave VR. Evaluation of multiple choice questions using item analysis tool: a study from a medical institute of Ahmedabad, Gujarat Int J Community Med Public Health 2017;4:1876-81.
 7. Item difficulty item discrimination. [Internet] [cited 2017 July 5]. Available from: <http://www.omet.pitt.edu/docs/OMET%20Test%20and%20Item%20Analysis.pdf>
 8. Varma S. Preliminary item statistics using point-biserial correlation and p-values. [Internet] [cited 2017 July 6]. Available from: http://www.eddata.com/resources/publications/EDS_Point_Biserial.pdf
 9. Bhoopatiraj C, Chellamani K. Analysis of test items on difficulty level and discrimination index in the test for research in education. IRJC 2013; 2 (2):189-93.
 10. Tarrant M, Ware J. Impact of item-writing flaws in multiple-choice questions on student achievement in high-stakes nursing assessments. Med Educ 2008; 42: 198-206.
 11. Mukherjee P, Lahiri SK. Analysis of Multiple Choice Questions (MCQs): Item and Test Statistics from an assessment in a medical college of Kolkata, West Bengal. IOSR Journal of Dental and Medical Sciences (IOSR-JDMS) e-ISSN: 2279-0853, p-ISSN: 2279-0861. Volume 14, Issue 12 Ver. VI (Dec. 2015), PP 47-52
 12. 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.
 13. 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 Ahmedabad, Gujarat. Indian J Community Med 2014; 39:17-20.
 14. Karelia BN, Pillai A, Vegada BN. The levels of difficulty and discrimination indices and relationship between them in four response type multiple choice questions of pharmacology summative tests of year II MBBS students. IeJSME 2013; 7(2): 41-6.
 15. Hingorjo MR, Jaleel F. Analysis of One-Best MCQs: the Difficulty Index, Discrimination Index and Distractor Efficiency. J Pak Med Assoc; Vol. 62, No. 2, February 2012: 142-147.