Original Research Article

Item analysis of multiple-choice questions - assessing the quality of MCQs being used to assess medical students

Pratima Gedam[1], Prateek Goyal[2*], Namita Srivastav[3]

[1]Assistant Professor, [2]III[rd] Year PG Resident, [3]Statistician Cum Lecturer, Department of Community Medicine, V.M.M.C. And Safdarjung Hospital, Ansari Nagar New Delhi, INDIA.
Email: pratimagedam2007@rediffmail.com, prateekgoyalkaku@gmail.com

Abstract

Background: MCQs are an objective method of assessing students in an exam. They help in eliminating the examiner's bias. However, MCQs should be carefully constructed for a good assessment. This study was done to analyze the quality of MCQs using item analysis. **Aim:** To analyze the Quality of MCQs being used by Community Medicine department to assess students in exam. **Materials and Methods:** Fifty MCQs were administered to 73 students and analysed on 3 parameters: Difficulty Index (DIF), Discrimination Index (DI) and Distracter Efficiency (DE). **Results:** Mean Difficulty Index was 47.0 ± 27.3, that is, our test had an acceptable level of Difficulty Index as a whole. Mean Discrimination Index was 0.24 ± 0.2 which means our test had an acceptable level of Discrimination Index as a whole. Mean Distracter Efficiency was 52.6 ± 23.6 which reveals that about half of the distracters in the whole test were functional, while the other half were non-functional. Difficulty Index was positively correlated with Discrimination Index and negatively correlated with Distracter efficiency. **Conclusion:** The majority of the items fulfilled the criteria of acceptable difficulty and discrimination, which means the MCQs used were of good quality. However, many distracters were non-functional, which needs improvement.

Keywords: Item analyses, Difficulty Index, Discrimination Index, Distracter efficiency, MCQs.

[*]**Address for Correspondence:**
Dr Prateek Goyal, III[rd] Year PG Resident, Department of Community Medicine, V.M.M.C. And Safdarjung Hospital, Ansari Nagar, New Delhi, INDIA.
Email: prateekgoyalkaku@gmail.com

Received Date: 17/05/2022 Revised Date: 12/09/2022 Accepted Date: 29/11/2022

Access this article online

Quick Response Code:

Website: www.medpulse.in

DOI: https://doi.org/10.26611/10112431

INTRODUCTION

Objective evaluation is a very important tool in the education system of today. It is an easier method as compared to subjective evaluation, but has its own strengths and limitations. MCQs or "items" are used all over the world to assess students. But for a good assessment, multiple choice questions (MCQ) should be well designed, which is a difficult and time-consuming process. A Well-constructed MCQ is a very useful tool for examination that can cover complete subject or any topic within it with objectivity across all levels of cognitive domain. It also decreases the evaluator's bias by minimizing examiner's judgment during scoring. But again, development of standardized MCQ is a time-consuming task, and if an MCQ is not constructed well, it can be easier or difficult for the students. If the options of the MCQ are not meeting the standards, it will lead to decreased recalling by students, less use of comprehension and/or problem-solving skills and will lead students towards guessing the correct answer.[1] Some instructors believe that MCQs are just multiple-guess items and that MCQs are capable of testing only factual information. So, they do not consider them suitable for testing higher-order cognitive skills. This may be true for ill-constructed MCQs/items but for well-constructed multiple-choice items, it is now accepted that "they test many higher

How to cite this article: Pratima Gedam, Prateek Goyal, Namita Srivastav. Item analysis of multiple-choice questions - assessing the quality of MCQs being used to assess medical students. *MedPulse International Journal of Community Medicine.* December 2022; 24(3): 14-17. https://www.medpulse.in/

cognitive skills of Blooms taxonomy such as knowledge, application, analysis and synthesis; which is a necessity for the assessment of health care professionals."[2]

Item analysis is a valuable, relatively simple but an effective process to check the reliability and validity of MCQs. This is helpful in three aspects. It assesses[1]:

1. How much the question is difficult or easy to attempt for the student, which is called the Difficulty Index (DIF).
2. To what index a question can discriminate the students having good knowledge about subject from the students having less knowledge of the subject, which is called the Discrimination index (DI).
3. To what extent are the wrong options able to distract the students from the correct answer, which is called the Distracter Efficiency (DE).

So, by running an item analysis, an examiner looks at the responses to individual items by the students, to assess the quality of questions/items and also to the quality of test as a whole. This is a very good approach to find out questions which are well constructed and can be kept in a question bank for future use. It also helps the examiner to improve his/her skill in constructing a good item/MCQ and to find out which topics of part of a chapter may not be clear to students and needs further clarifications. It also provides teachers with feedback, which may help them to improve their standards of teaching. It also helps to find out question which are extremely easy or extremely tough, or which have an error within them, thus helping the examiner to make a decision whether to keep or remove such questions for further tests.[3] With this background, this study was done with the objectives of this study being analyzing the quality of MCQs used in Community Medicine department of VardhmanMahavir Medical College and Safdarjung Hospital, New Delhi and to determine the relationship between Difficulty index, Discrimination index and distracter efficiency.

MATERIALS AND METHODOLOGY

Study Design: Cross-sectional study

Study Duration: 1 month

***Study Setting*:** Department of Community Medicine of Vardhman Mahavir Medical College and Safdarjung Hospital.

***Study Population*:** 3rd year Medical students

***Inclusion Criteria*:**3rdyear M.B.B.S.students studying in the batch for the period of one month.

***Exclusion Criteria*:** Students who were absent for test.

Sample Size: 73 students attempted 50 M.C.Q.s

***Sampling Technique*:** Complete enumeration

Study Tool: A set of carefully constructed 50 questions administered to students as a test (to be completed within 60 minutes). All M.C.Q.S were single based response type with one answer and three distractors each. Each correct response was given 1 mark while there was no negative marking. M.C.Q.s were framed by faculty in the department of community Medicine and were validated by H.O.D. of the department of Community Medicine.

***Method*:** After the students submitted the test, the responses were evaluated. The first steps for item analysis was to arrange all the students in the descending order according to their scores in the test. Then, Top 27% of the students were taken as high achievers (H) and bottom 27% as (L) as low achievers. Three indices were calculated for each M.C.Q. and calculations were made using the following formulae[3]:

Difficulty index (DIF) = (H + L)/n × 100.

Discrimination index (DI) = 2 X (H – L)/n

Distracter Efficiency (DE) = FD/TD

Where;

H = number of students answering correctly in high achievers' group.

L = number of students answering correctly in the low achiever's group.

n = total number of students in both groups including non-responders.

FD = Functional Distracter (opted by >5% of students)

NFD = Nonfunctional Distracter (Opted by <5% of students)

TD = Total Distracters

Interpretation [3]

Difficulty index (DIF)

DIF≤ 30%	-	Difficult
DIF= 30-70%	-	Acceptable
DIF ≥ 70%	-	Easy

So if the DIF is high, the question is easier.

Discrimination index (DI):

DI = Negative	-	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 ≥ 0.4	-	Excellent discrimination

Data analysis and statistical methods

The data was entered in CITAS-2016 software and was cleaned for errors and missing values. Data analysis was done using CITAS-2016and licensed SPSS software version 21.0. The data was reported in percentages and Mean ± S.D. The relationship between Difficulty Index, Discrimination Index and Distracter Efficiency was determined using Pearson correlation analysis. $P < 0.05$ was taken as significant.

RESULTS

Total 73 students took the test. On the basis of marks obtained, top 27% of these students (i.e. 20 students) were labeled as High achievers and bottom 27% of these students were labeled as low achievers. Mean Difficulty Index was 47.0 ± 27.3. That is, our test had an acceptable level of Difficulty Index as a whole. Table 1 shows that out of 50 Question, 12 were easy, 23 had an acceptable level of difficulty Index and 15 questions were Difficult. ***[Table 1 near here]***

Table 1: Difficulty Index of the questions in the test (N=50)

Difficulty Index	Category	No. of Questions (N=50)
DIF≤ 30%	Difficult	15
DIF= 30-70%	Acceptable	23
DIF ≥ 70%	Easy	12

Mean Discrimination Index was 0.24 ± 0.2. That is, our test had an acceptable level of Discrimination Index as a whole. Table 2 shows that out of 50 Question, only 2 questions had a Negative DI, 19 questions had a Poor DI, 8 questions had an Acceptable DI, 11 questions had a Good DI and 10 questions had an Excellent DI. ***[Table 2 near here]***

Table 2: Discrimination Index of the questions in the test (N=50)

Discrimination index	Category	No. of Questions (N=50)
DI = Negative	Defective item/wrong key	2
DI = 0-0.19	Poor discrimination	19
DI = 0.2 – 0.29	Acceptable discrimination	8
DI = 0.3 – 0.39	Good discrimination	11
DI ≥ 0.4	Excellent discrimination	10

Mean Distracter Efficiency was 52.6 ± 23.6. That is, about half of the distracters in the whole test were functional, while the other half were nonfunctional distracters. Table 3 shows that out of 50 questions, there were 6 questions with all 3 distracters as functional, 17 questions with 2 out of 3 distracters functional and 27 questions with 1 out of 3 distracters functional. None of the questions had zero functional distracters. ***[Table 3 near here]***

Table 3: Distracter Efficiency of the questions in the test (N=50)

Distracter Efficiency	Category	No. of Questions (N=50)
DE = 0%	(FD-0, NFD-3)	0
DE = 33%	(FD-1, NFD-2)	27
DE = 67%	(FD-2, NFD-1)	17
DE = 100%	(FD-3, NFD-0)	6

Correlations

There was a negative and non-significant association between Discrimination Index and Distracter Efficiency. Pearson correlation between Difficulty Index and Distracter Efficiency showed that DIF is negatively correlated with DE (r = -0.316). The association is significant. Pearson correlation between Difficulty Index and Discrimination Index showed that DIF is positively correlated with DI (r=0.285). The association is significant. ***[Table 4 near here]***

Table 4: Correlation between DIF, DI and DE

		DIF	DIS	DE
DIF	Pearson Correlation	1	**0.285**	**-0.316**
	Sig. (2- tailed)		0.044	0.025
	N	50	50	50
DI	Pearson Correlation	**0.285**	1	-0.127
	Sig. (2- tailed)	0.044		0.379
	N	50	50	50
DE	Pearson Correlation	**-.316**	-0.127	1
	Sig. (2- tailed)	0.025	0.379	
	N	50	50	50

Figure 1 shows linear, quadratic and cubic regression between Discrimination Index and Difficulty Index. Quadratic and Cubic Regression models reveals that initially, with an increase in Difficulty Index, Discrimination Index also increases. It reaches a peak, and then starts falling with further increase in Discrimination Index. ***[Figure 1 near here]***

Figure 1 Linear, quadratic and cubic regression between Discrimination Index and Difficulty Index.

DISCUSSION

MCQs are a very convenient tool for assessing the knowledge and competence of a student. The use of MCQs also eliminates examiners bias. An MCQ must be carefully designed to assess the factual recall capacity, comprehension and understanding of a medical student. A test as a whole should neither be too easy, nor too difficult. Also, one of the motives behind any test is to differentiate between students with good and bad performance. Doing an Item analysis with Difficulty Index, Discrimination Index and Distracter Efficiency helps us to know about the validity and reliability of a test, and provides further scope for improvement.[4] Mean Difficulty Index of 50 questions of the test was 47.0 ± 27.3, which was well within the acceptable range of difficulty. Too difficult items may decrease the moral of students while too easy items will lead to students being overconfident. Difficulty Index of tests/exams reported by many studies (Table 5) usually lie in the acceptable range of difficulty. Mean Discrimination Index was 0.24 ± 0.2. Our test had an acceptable level of Discrimination Index, but it has scope for further improvement. From Table 5, we see that Discrimination Index of tests/exams usually has an acceptable to good level of Discrimination. Mean Distracter Efficiency was 52.6 ± 23.6. That is, about half of the distracters in the whole test were functional. Also, all the questions in the test had at least 1 Functional Distracter. But, From Table 5, We see that Distracter efficiency is usually on the higher

side as compared our study. These Non functioning Distracters can be replaced with a more plausible option to increase the distracter efficiency, but this may also lead to a fall in the level of difficulty index, making the questions more difficult.

Table 5: DIF, DI and DE as reported by various studies, compared to the current study.

	DIF	DI	DE
Current Study	47.0 ± 27.3	0.24 ± 0.2	52.6 ± 23.6
Majhabeen *et al.*,2017	58.74 ± 14.39	0.35 ± 0.16	63.55 ± 27.47
Mukharjee *et al.*,2015	61.92 ± 25.1	0.31 ± 0.27	-
Shete *et al.*,2015	54.0 ± 26.0	0.21 ± 0.14	-
Menon *et al.*2017	44.8 ± 17.13	0 .37 ± 0.18	87.5 ± 17.2
Rao *et al.*,2016	50.16 ± 16.15	0.34 ± 0.17	89.99 ± 24.42
Pande SS *et al.*,2013	52.53 + 20.59	0.30+ 0.18	-

Pearson correlation between Difficulty Index and Distracter Efficiency showed that DIF is negatively correlated with DE. This implies that when a question has more functional distracters, the question becomes more difficult. Quadratic and Cubic Regression models between Discrimination Index and Difficulty Index reveal a dome shaped curve. This implies that question with high DIF (very easy) as well as questions with low DIF (very difficult) questions, both have a poor Discrimination Index. Similar results were found by Shete *et al.*, who found that Difficulty Index and Discrimination index are positively correlated and the maximum discrimination is for the items with moderately easy or difficult items.[3] Menon *et al.* also pointed out that the relationship between Difficulty Index and Discrimination Index is often dome shaped.[4]

CONCLUSION

In the present study, the majority items fulfilled the criteria of acceptable difficulty and discrimination, which means the MCQs selected were of good quality. Moderately easy/difficult had maximum discrimination ability. Very easy and very difficult items displayed poor discrimination. Items with negative and poor discrimination will be reviewed, reconstructed, and added to the departmental MCQ bank. The results of this study should initiate a change in the way MCQ test items are selected for any examination, and there should be proper assessment strategy as part of the curriculum development. Much more of these kinds of analysis should be carried out after each examination to identify the areas of potential weakness in the one best answer type of MCQ tests to improve the standard of assessment.

Acknowledgement

I would like to thank all the respected Faculty members with whose inputs and suggestions this 50-Question based test was created to assess the knowledge of 3rd Year students.

Ethical consideration

The study did not involve any method that put the subjects or the investigators at risk. There is no conflict of interest. The data collected and generated was kept safe in a password locked computer of the principal investigator and the responses of any student were not shared with other. However, for academic purposes, students weres known their own total marks obtained in the test.

REFERENCES

1. Mahjabeen W, Alam S, Hassan U, Zafar T, Butt R, Konain S, Rizvi M. Difficulty index, discrimination index and distractor efficiency in multiple choice questions. Annals of PIMS-Shaheed Zulfiqar Ali Bhutto Medical University. 2017;13(4):310-5.
2. 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 J Dent Med Sci. 2015;1:47-52.
3. Shete AN, Kausar A, Lakhkar K, Khan ST. Item analysis: An evaluation of multiple choice questions in Physiology examination. J Contemp Med Edu. 2015;3(3):106-9.
4. Menon AR, Kannambra PN. Item analysis to identify quality multiple choice questions. National J. Lab. y Medi. 2017;6(2):M007-10
5. Rao C, Prasad HK, Sajitha K, Permi H, Shetty J. Item analysis of multiple choice questions: Assessing an assessment tool in medical students. International Journal of Educational and Psychological Researches. 2016 Oct 1;2(4):201.
6. Pande SS, Pande SR, Parate VR, Nikam AP, Agrekar SH. Correlation between difficulty and discrimination indices of MCQs in formative exam in physiology. South-East Asian Journal of Medical Education. 2013 Jun 30;7(1):45-50.
7.

Source of Support: None Declared
Conflict of Interest: None Declared