

QUALITY ANALYSIS OF MULTIPLE CHOICE QUESTIONS

¹BENISH MEHMOOD, ²SEHAR SULTAN, ³MUHAMMAD AZHAR, ⁴ASMA SHAKOOR,
⁵NAUMAN BARI KHAN, ⁶ZAINAB NIAZI

ABSTRACT

The most well-known and widely accepted method to assess knowledge and professional competencies of medical students are multiple choice questions. They are subjectable to automated computer marking, minimizing the chance of inter examiner variability in marking. Construction of good quality MCQ not only requires dedication but also familiarity with MCQ items writing guidelines and willingness towards change in personal writing habits. It is therefore necessary to frequently assess quality of MCQs. The aim of this study was to evaluate the quality of multiple choice question (MCQ) in Community Dentistry Undergraduate Examination.

A total of 25 one best type MCQs or items with four options were used for assessment. Post-validation was done by determining difficulty Index (DIF I/P), Discrimination Index (DI) and Distractor efficiency (DE). The test had an acceptable difficulty level with $69.76\% \pm 18.69\%$ mean difficulty index. Discrimination index of the test was also satisfactory with 0.23 ± 0.15 mean point biserial correlation coefficient. Out of a total of 75 distractors, 64 were functional hence distractor efficiency was $85.33\% \pm 21.69\%$. The reliability coefficient (KR20) of the test was 0.24. 56% of MCQs had adequate difficulty and discrimination indices. There were 40% MCQs that were too easy while 4% were too difficult. 64% of MCQs had 100% distractor efficiency

Item analysis of multiple choice questions can be valuable as it enables examiner to identify the shortcomings or any flaw encountered during item construction.

Key Words: Community Dentistry, undergraduate, knowledge, professional competencies, MCQs

This article may be cited as: Mehmood B, Sultan S, Azhar M, Shakoor A, Khan NB, Niazi Z. Quality Analysis of Multiple Choice Questions. Pak Oral Dent J 2021; 40(4):236-39.

INTRODUCTION

Assessment has a pivotal role in education as it encourages learning. It helps in highlighting the pros and cons of training programmes and also identification of strengths and shortcomings of candidates. A well planned assessment has very powerful impact on

learning and curriculum. Different researches have shown a positive correlation between assessment tools and examinees' choice of learning approach.¹⁻³ Quality of questions also has a significant effect on learning methodology. When examiner focuses more on recall of isolated facts, this would encourage superficial learning approach.⁴

Careful attention is required to select a feasible, valid and reliable assessment tool, to distinguish among satisfactory and unsatisfactory performers.⁵ Several assessment tools exist to assess learning and performance like short essay questions, short answer questions, modified essay questions, EMQs, objective structured clinical examination, multiple choice questions etc. Literature shows that essay type questions are more useful tool for testing recall of knowledge³ whereas essay type are more appropriate for undergraduate assessment.⁶

The most well-known and widely accepted method to assess knowledge and professional competencies of undergraduate and post graduate medical students

¹ Dr Benish Mehmood, BDS, MDPH, Associate Professor, Department of Community Dentistry, Army Medical College, National University of Medical Sciences, Rawalpindi E-mail: beenishmehmood@hotmail.co.uk Cell: 0322-5123027

² Dr Sehar Sultan, BDS, MPH, Assistant Professor, Department of Community Dentistry, Army Medical College, National University of Medical Sciences, Rawalpindi.

³ Dr Muhammad Azhar, BDS, Demonstrator, Department of Oral Biology, Army Medical College, National University of Medical Sciences, Rawalpindi.

⁴ Dr Asma Shakoor BDS, MSc Public Health (U.K), MFDS RCSEd, Associate Professor, Community Dentistry Department, Institute of Dentistry, CMH Lahore Medical College, Lahore.

⁵ Dr Nauman Bari Khan, BDS, MSc, Assistant Professor, Department of Oral Biology, Army Medical College, National University of Medical Sciences, Rawalpindi.

⁶ Dr Zainab Niazi, BDS, Mphil, Assistant Professor, Oral Pathology, Islamabad Dental college, Islamabad.

Received for Publication: Feb 12, 2020

Revised: Apr 28, 2020

Approved: May 6, 2020

are MCQs.^{4,7} Multiple choice questions can be used individually or with other tools for assessment. Use of MCQs as assessment tool is very advantageous as it not only evaluates medical information but also assess the advanced level of critical thinking.⁸⁻¹⁰ They are subjectable to automated computer marking, minimizing the chance of inter examiner variability in marking. Odd pattern of results can be highlighted easily through psychometric analysis.

Construction of good quality MCQ not only requires dedication but also familiarity with MCQ items writing guidelines and willingness towards change in personal writing habits.^{11,12} It is therefore necessary to assess quality of MCQs through "item analysis", in terms of Discrimination Index, Difficulty Index and Distractor efficiency.¹³ The item discrimination index helps to identify how well an item is able to differentiate between knowledgeable and those who are not. The skill levels is estimated through difficulty Index and is assessed through proportion of individuals passing an item. Greater the value of difficulty index, easier the item will consider and vice versa. The distractor analysis helps to check effectiveness of the incorrect options in determining quality of an MCQs.¹⁴ A study was conducted in Pakistan to check the cognitive level of both MCQs and SEQs, reported that 60% of the MCQs and 83.33% of SEQs were of recall level.⁴

The aim of this study was to evaluate the quality of multiple choice question (MCQ) items in Community Dentistry Undergraduate Examination by determining difficulty Index (DIF I/P), Discrimination Index (DI) and Distractor efficiency (DE). Information thus obtained from this study will help us not only to improve the quality of MCQs but also to assess student's professional competencies in a better way.

METHODOLOGY

A total of 50 undergraduate Dental students of second year BDS were assessed in subject of Community dentistry in July 2019. A total of 25 one best type MCQs or items with four options were used for assessment. Students were asked to select one best answer from these four choices. A correct item was given one mark and zero to the wrong one, with no negative marking. Post-validation was done by item analysis. SPSS version 24 was used for data analysis.

Difficulty Index or P value was calculated by using the formula:

$$\text{DIF I or P} = (c/n) \times 100$$

Where "c" is the number of correct responses for each item and "n" is the total number of students. A high difficulty index indicated that item was easy while a low index indicated that the item was difficult.

Discrimination index or d value index helps to identify how well an item is able to differentiate between knowledgeable and those who are not and it was calculated using the point biserial correlation coefficient also known as point biserial index (PBI).

Distractor efficiency was determined on basis of number of MCQs with non-functional distractors (NFD) (option selected by <5% of students) in it. Mean and standard deviation for difficulty Index (DIF I/P), Discrimination Index (DI) and Distractor efficiency (DE), were calculated. Percentage of items falling in different categories of DI and DIF I/P were also calculated.

RESULTS

A test with 25 multiple choice questions of choose the best answer type options was answered by 50 undergraduate dental students. Score "1" was given for each correct response and "0" for each incorrect answer. The mean score was 17.44 ± 2.4 with a range of 11-22.

The test had an acceptable difficulty level with $69.76\% \pm 18.69\%$ mean difficulty index. 14 (56%) MCQs were too easy having a difficulty index above 70%. There were 10 (40%) questions in the acceptable range of difficulty index i.e. 30-70%. Of these 10 questions, 3 (12%) had a good difficulty index ranging between 50-60%. Only 1 out of 25 items (4%) was too difficult with 26% difficulty index (Fig. 1).

Discrimination index of the test was satisfactory with 0.23 ± 0.15 mean point biserial correlation coefficient. There were 11 (44%) questions, that had poor point biserial index (<0.2) (Fig. 1), 6 (24%) questions had a fair point biserial index (0.2-0.29), 3 (12%) items had a good point biserial index (PBI) (0.3-0.39) and 5 (20%) questions had an excellent PBI (0.4-0.7). Of the 11 questions having poor PBI, 1 item had a negative PBI (-0.09).

Out of a total of 75 distractors, 64 were functional hence distractor efficiency was $85.33\% \pm 21.69\%$. There were 16 (64%) items that had 100% distractor efficiency, 7 (28%) had one non-functional distractor while 2 (8%) MCQs had 2 non-functional distractors in each item (Fig. 1). No MCQ had 3 non-functional distractors.

The reliability coefficient (KR20) of the test was 0.24. 56% of MCQs had adequate difficulty and discrimination indices. 40% MCQs were too easy while 4% were too difficult. 44% MCQs had a low discrimination index. 64% of MCQs had 100% distractor efficiency, 28% had one non-functional distractor and 8% items had 2 non-functional distractors each.

DISCUSSION

Multiple choice questions are the most frequently used assessment tool for assessing cognitive domain of

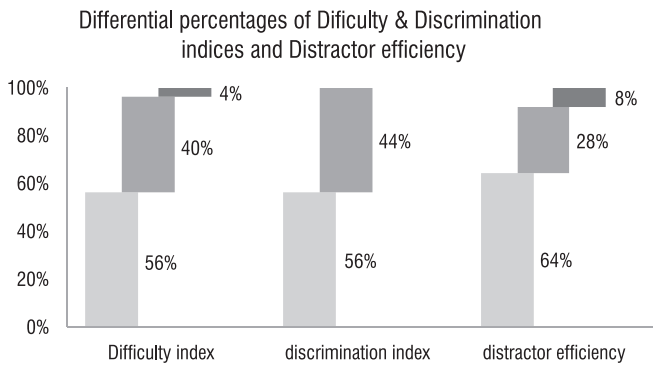


Fig 1: Percentage of Mcqs Having Various Levels of Difficulty and Discrimination Indices and Distractor Efficiencies.

learner, whereas the affective and psychomotor skills cannot be evaluated.¹⁵ MCQs are considered as quickest tool to assess large number of students in a short span of time, are easy to mark valid, feasible and reliable.

Creating a well-constructed MCQ can be a challenging and time consuming task and needs to be evaluated for quality. One such valuable tool used to test the reliability & validity of MCQs is Item analysis.¹⁶ It helps examiner to determine whether to keep the item in the test, review it or discard. Commonly used parameters for item analysis are discrimination index (DI) or point biserial, difficulty index (DIF I) and distractor efficiency (DE).

A study conducted by Patel and Mahajan on item analysis of MCQs of 150 MBBS students reported that 40 (80%) items out of total 50, were in acceptable range ($P = 30-70\%$), whereas 10 (20%) items were in unacceptable range ($P < 30\%$ or $P > 70\%$).¹⁷ Another study by Patil and Patil, for 100 items on 100 MBBS students reported mean DIF I of 48.90 ± 13.72 . 25 (25%) items out of 100 were ideal (50–60%), 22 (35%) items were too difficult ($P < 30\%$) whereas 18 (18%) items were too easy ($P > 70\%$).¹⁸ Mehta and Mokhasi, in their study on item analysis of 50 anatomy MCQS, found a mean DIF I of 63.06 ± 18.95 with DIF I of 31 (62%) items in the acceptable range ($P = 30-70\%$), 16 (32%) items were too easy and 3 (6%) items were too difficult.¹⁹ In this study difficulty level was acceptable with $69.76\% \pm 18.69\%$ mean difficulty index. There were 14 (56%) MCQs that were too easy having a difficulty index above 70%. Ten (40%) questions were in the acceptable range of difficulty index i.e. 30-70%. Of these 10 questions, 3 (12%) had a good difficulty index ranging between 50-60%. Only 1 out of 25 items (4%) was too difficult with 26% difficulty index. 14 too easy Items were revised and kept for subsequent use along with items within acceptable range.

Another important parameter of item analysis used for discrimination among high achievers and low

achievers is Discrimination index (DI). It has range from 0 and 1, where a greater value shows increased ability of MCQ to discriminate between a high achieving student and a low achieving student. Sometimes the value can be negative indicating a flaw in item and it is called negative DI. This negative value shows that more number of low achievers correctly answered the question as compare to high achievers may be due to wrongly marked answer key or ambiguous questions.

In our study, Discrimination index of the test was satisfactory with 0.23 ± 0.15 mean point biserial correlation coefficient. There were 11 (44%) questions, that had poor point biserial index (< 0.2), 6 (24%) questions had a fair PBI (0.2-0.29), 3 (12%) items had a good PBI (0.3-0.39) and 5 (20%) questions had an excellent PBI (0.4-0.7). Of the 11 questions having poor PBI, only 1 item had a negative PBI (-0.09). Study conducted on 20 MCQs by Singh *et al.*, reported 20% items with $DI \geq 0.20$ & ≤ 0.35 while 30% items had < 0.2 DI.²⁰ Patel and Mahajan, in their study on 50 MCQs items reported $DI \geq 0.20$, and ≤ 0.35 for 21 items.¹⁷ Study by Mehta and Mokhasi, showed mean DI of 0.33 ± 0.18 .¹⁹ Another study on item analysis of MCQs in subject of Prosthodontics, the mean DI was 0.12 ± 0.13 with 67.5% of the test items having a $DI < 0.2$.²¹ Study on item analysis at Bahrain found mean discrimination index ranged from 0.20–0.34.²²

A quality item should consist of plausible distractors. Mean distractor efficiency (DE) in the present study was $85.33\% \pm 21.69\%$. 16 (64%) items had 100% distractor efficiency, 7 (28%) had one non-functional distractor while 2 (8%) MCQs had 2 non-functional distractors in each item. Similar results were reported by Hingorjo *et al.*, reported a mean DE of 81.4%¹⁴ while Gajjar *et al.*, reported a mean DE of 88.6 ± 18.6 showing good efficiency of distractors.¹³

This is the first study on item analysis in subject of public health dentistry. The study has limitations as items are not in greater number. Further research is therefore recommended for further improvement in preparation of quality MCQs.

CONCLUSION

Post hoc analysis of multiple choice questions can be valuable as it enables examiner to identify the shortcomings or any flaw encountered during item construction. This not only strengthens the MCQ bank but also a good percentage of valid and reliable items get added for assessment. Training sessions of faculty members are very important as they help educators in improving the quality of their MCQs.

REFERENCES

- 1 Trigwell K, Prosser M. Improving the quality of student learning: the influence of learning context and student approaches to

- learning on learning outcomes. *High Educ.* 1991;22(3):251–66
- 2 Biggs J, Tang C. *Teaching for Quality Learning at University*. 4th ed. Philadelphia, USA: Open University Press; 2011. p. 103.
 - 3 Reid WA, Duvall E, Evans P. Relationship between assessment results and approaches to learning and studying in year two medical students. *Med Educ.* 2007 Aug;41(8):754-62
 - 4 Baig M, Ali SK, Ali S, Huda N. Evaluation of Multiple Choice and Short Essay Question items in Basic Medical Sciences. *Med Educ.* 2007 Aug;41(8):754-62
 - 5 Peck C. *Principles of Sound Assessment Practice in Health Professions Education*. *EC Psychology and Psychiatry.* 2017;5(5):150-57.
 - 6 Oyebola DO, Adewoye OE, Iyaniwura JO, Alada RA, Fasanmade AA, Raji YA. Comparative study of students' performance in preclinical physiology assessed by multiple choice and short essay questions. *Afr J Med Med Sci.* 2000;29(3-4):201-5
 - 7 Kemp JE, Morrison GR, Ross SM. *Developing evaluation instruments. Designing Effective Instruction*. New York: MacMillan College Publishing Company; 1994. pp. 180–213.
 - 8 Coderre SP, Harasym P, Mandin H, Fick G. The impact of two multiple-choice question formats on the problem-solving strategies used by novices and experts. *BMC Med Educ.* 2004;5(4):23
 - 9 Bauer D, Holzer M, Kopp V, Fischer MR. Pick-N multiple choice-exams: a comparison of scoring algorithms. *Adv Health Sci Educ Theory Pract* 2011; 16: 211–21.
 - 10 Vanderbilt AA, Feldman M, Wood IK. Assessment in undergraduate medical education: a review of course exams. *Med Educ Online.* 2013 Mar 6;18(1):1-5. doi: 10.3402/meo.v18i0.20438
 - 11 Abdulghani HM, Ahmad F, Irshad M, Khalil MS, Al-Shaikh GK. Faculty development programs improve the quality of multiple choice questions items' writing. *Sci Rep* 2015; 5: 9556
 - 12 AlFaris E, Naeem N, Irfan F, Qureshi R, Saad H, AlSadhan R. A One-day dental faculty workshop in writing multiple choice questions: an impact evaluation. *J Dent Educ.* 2015 Nov;79(11):1305-13.
 - 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 of Ahmedabad, Gujarat. *Indian J Community Med.* 2014;39(1):17-20.
 - 14 Hingorjo MR, Jaleel F. Analysis of One-Best MCQs: the Difficulty Index, Discrimination Index and Distractor Efficiency. *J Pak Med Assoc.* 2012;62(2):142-47.
 - 15 Singh T, Anshu . *Principles of Assessment in Medical Education*. New Delhi: Jaypee Brothers Medical Publishers; 2012. p. 89.
 - 16 Considine J, Botti M, Thomas S. Design, format, validity and reliability of multiple choice questions for use in nursing research and education. *Collegian.* 2005 Jan;12(1):19-24
 - 17 Patel KA, Mahajan NR. Itemized analysis of questions of multiple choice question exam. *Int J Sci Res.* 2013;2(2):279–80.
 - 18 Patil VC, Patil HV. Item analysis of medicine multiple choice questions (MCQs) for under graduate (3rd year MBBS) students. *Res J Pharm Biol Chem Sci.* 2015;6(3):1242–51.
 - 19 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.
 - 20 Singh JP, Kariwal P, Gupta SB, Shrotriya VP. Improving multiple choice questions through item analysis: an assessment of the assessment tool. *Int J Basic Sci Appl Res.* 2014;1(2):53-57.
 - 21 Rehman A, Aslam A, Hassan SH, Item analysis of multiple choice questions. *Pak Oral Dental J.* 2018;38(2):291-93
 - 22 Kheyami D, Jaradat A, Al-Shibani T, Ali FA. Item Analysis of Multiple Choice Questions at the Department of Paediatrics, Arabian Gulf University, Manama, Bahrain. *Sultan Qaboos Univ Med J.* 2018;18(1): 68–74.

CONTRIBUTIONS BY AUTHORS

- | | |
|-------------------|---|
| 1 Benish Mehmood: | Study concept and design, Introduction, Discussion writing, |
| 2 Sehar Sultan: | Data collection, Methodology |
| 3 Muhammad Azhar: | Data analysis, Results |
| 4 Asma Shakoor: | Critical revision |