SHORT COMMUNICATION

Quality Assurance of Multiple-Choice Questions Test Through Item Analysis

Mashhood-uz-Zafar Farooq[1*], Shama Mashhood[2]

ABSTRACT

Objective: To analyze the value of in use multiple choice questions of ophthalmology by finding discrimination index, distractor efficiency and difficulty index.

Study Design: Cross-sectional study.

Place and Duration of Study: The Study was conducted at the Department of Ophthalmology, Karachi Institute of Medical Sciences, Karachi, Pakistan from January 2022 to March 2022.

Methods: The study included result of all multiple-choice questions administered during pre-annual examination of the year 2021. There were 45 multiple-choice questions with one correct option and three distractors. Analysis of each item was performed to find difficulty discrimination index and distractor efficiency. Data were entered and analyzed by SPSS software 20.0. Frequency and percentage were calculated for all categorical variables and mean and standard deviation were considered for all continuous variables. Difficulty and discrimination index and distractor efficiency were calculated for multiple-choice questions.

Results: Overall, 68.9% had good/acceptable levels of difficulty and were stored, whereas 24.4% were too easy and 6.7% were too complex and confusing. Discrimination analysis demonstrated 27 items to be excellent, 13 good and 5 having poor discrimination. Distractor efficacy was found to be 93.32±19.60. Out of 135 distractors, 129 were functional while non-functioning distractor (NFD) were 6.

Conclusion: Item analysis is a valuable assessment tool that identifies better multiple-choice questions to be retained while discarding or reviewing the weak ones. Faculty development programs should be organized for improving item writing skills of faculty.

Keywords: *Difficulty Index, Discrimination Index, Faculty Development Program, Item Analysis, Item Writing Skills.*

How to cite this: *Farooq MZ, Mashhood S. Quality Assurance of Multiple Choice Questions Test Through Item Analysis. Life and Science. 2023; 4(4): 501-507. doi: http://doi.org/10.37185/LnS.1.1.315*

Introduction

Medical education has undergone tremendous changes in the recent past. Traditional curriculum has shifted from subject to system-based. The current changes in curricula have resulted in significant improvements in student overall grade.[1] Teaching is now more interactive than solely lecture based. Problem based learning has been incorporated in curricula. In response to changes in teaching and learning process, assessment has also been improved. Assessment is a process of data collection to find the strength and weaknesses of student learning and thus has a crucial role in learning process.[2] Therefore assessment not only provides student score, it also provide valued evidence about student learning and growth.[3] The assessment is a tool that besides identifying barriers in learning and understanding, also helps teachers in improving teaching method and identifying the gaps and has a positive influence on student learning as well as teaching.[4] However, the objectives of assessment could only be achieved when

[1]*Department of Ophthalmology*
Karachi Institute of Medical Sciences, Karachi, Pakistan
[2]*Department of Medical Education*
Karachi Medical and Dental College, Karachi, Pakistan

Correspondence:
Dr. Mashhood-uz-Zafar Farooq
Professor, Ophthalmology
Karachi Institute of Medical Sciences, Karachi, Pakistan
E-mail: drmzafarfarooq@gmail.com

Funding Source: NIL; Conflict of Interest: NIL
Received: Nov 22, 2022; Revised: Aug 24, 2023
Accepted: Sep 03, 2023

assessment tool is valid and reliable. In line with the change in curricula, assessment by traditional long question and answers is largely replaced by one best multiple choice questions (MCQs). Educational objectives are described by Bloom`s Taxonomy as a framework consisting of Knowledge, Concept, Application, Analysis and Evaluation.[5] Therefore well-organized MCQs should assess all teaching objectives of Bloom's Taxonomy. Examining the student performance to each MCQ refers to Item analysis. This, in turn, provides examination validity and reliability. In-depth statistical analysis of each MCQ is required for a decision to keep, review, or discard it.[6] Discrimination index (DI), difficulty index (DIFI), and distractor efficacy (DE) are commonly used parameters for item analysis.[7] DIFI refers to the percentage of correct responses out of total responses with values from 0 to 1, and if the items are easy, DIFI is high and difficult items represent low DIFI.[8] DI identifies high and low achievers with range from -1 to +1.[9] Items having value of DI>0.35 are acceptable and retained whereas DI value from 0.2 to 0.35 are acceptable but requiring revision and DI <0.2 are poor discriminator and discarded.[10] Generally, items with DI >0.2 are required to be retained. A positive correlation is found between item discrimination and difficulty indexes and an MCQ with moderate DIFI shows better DI.[11] However, highly difficult items may show negative DI.[12] DE determines the effectiveness of each option to be labelled as a functional distractor or non-functional distractor (NFD). Students are attracted to NFD if they have some deficient knowledge. Construction of the distractor is such that it is close to correct answer. If an MCQ has many NFDs, it becomes easier for a less knowledgeable student to choose correct answer thereby lowering discrimination ability.[10] Using functional distractors significantly improves the test quality while constructing a MCQ.[13] Type A MCQs are commonly used for assessment which consist of a stem with four to five options.[14] However, in terms of psychometric merits, five options are not superior to four option MCQs.[12]

In view of the extensive use of MCQs in all undergraduate and postgraduate medical examinations, we designed this study to determine the value of MCQs being used in ophthalmology.

Methods

The study was conducted at the Ophthalmology Department of Karachi Institute of Medical Sciences, Karachi, Pakistan from January 2022 to March 2022. Approval of this study was granted by the college's principal and allowed access to the examination data. The study was conducted as per the principles of the Helsinki Declaration of 1975, as revised in 2000. The identity of students was kept confidential who took the examination. There was no human participation in this study; therefore, formal ethical approval was not taken. The study included all MCQs items of pre-annual examination of the year 2021. The examination is conducted by the examination cell of the institute. The paper consisted of 45 MCQs that covered essential knowledge and skills. MCQs were pre-validated by the departmental colleagues.

A total of 88 students took the exam. The pre-annual examination is summative, and MCQs were not previously shared with the students. The students marked their answers in a separate answer sheet. Each MCQ had a lead-in followed by four options having one correct and three distractors. One mark was awarded for a correct answer and zero for a wrong one with no negative marking. The data consisted of MCQs, and student responses were entered and analyzed using Microsoft Excel Sheet 2013. Marks obtained were arranged in descending order. Marks were grouped in upper one third, middle one-third, and lower one third and termed as high achiever, middle achiever, and low achiever, respectively. DIFI, DI, and DE were analyzed. DIFI is used to identify the difficulty of the item and is calculated by the formula P=H+L x 100/N.[15] P is DIFI, H is the number of students in the high achiever group who answered correctly, L is the number of students in the low achiever group who also answered correctly, and N is the number of students in both groups, including the non-responders. It ranges from 0% to 100%. The result <30 is interpreted as too difficult and >70 as too easy, and these items are required to be revised or discarded, while the result between 30-70 is regarded as good and item retained.[16] DI is determined by the formula DI=2x{(H-L)/N}.[17] The number of students in the high achiever group giving correct answers are denoted as H, while L is students in the low achiever group giving correct

answer, and N refers to a total number of students in both groups, including the non-responders. DI reflects the discriminating power of an item between low and high achievers. An item <0.20 is regarded as a poor discriminator and is rejected, having a value from 0.20-0.35 as good but requiring revision and >0.35 as excellent and retained.[15] DE refers to an item having a functional or non-functioning distractor (NFD). An NFD is chosen by <5% of students.[16] When there is no NFD in a MCQ, DE is considered 100%, with 1 non-functioning distractor as 66.6%, with 2 NFD 33.33% and 0% with 3 NFD.[18] Items having 0 and 1 non-functioning distractor are stored while item having 2 needs revision and item having 3 are discarded.[15]

Statistical analysis

Statistical package for social sciences software 20.0 (SPSS) was used for analysis. Frequency and percentage were calculated for all categorical variables, and mean, and SD was considered for all continuous variables. DI, DIFI, and DE were calculated for MCQs.

Results

A descriptive analysis of MCQs was performed.

Table 1: Descriptive analysis of 45 one-best answer MCQs

Item description	Values
Total number of students	89
Total MCQs	45
Minimum score obtained for correct answer	12
Maximum score obtained for correct answer	41
Total mean score achieved	34.93±3.72
Mean score of higher achieving students	18.1667±2.94
Mean score of low achieving students	26.56±7.50
Mean of difficulty Index (DIF)	58.66±16.71
Mean of discrimination Index (DI)	0.37±0.17
Distractor efficacy (DE)	93.32±19.60

Table 2: Difficulty index and discrimination index of the 45 one-best MCQ

Question	DIFI	DI	Question	DIFI	DI	Question	DIFI	DI
1	75.0	0.23	16	53.0	0.60	31	33.0	0.06
2	81.0	0.23	17	76.0	0.46	32	60.0	0.73
3	40.0	-0.13	18	65.0	0.43	33	68.0	0.43
4	40.0	0.33	19	78.0	0.30	34	66.0	0.40
5	70.0	0.40	20	80.0	0.33	35	86.0	0.26
6	71.0	0.23	21	56.0	0.46	36	60.0	0.53
7	61.0	0.43	22	73.0	0.46	37	48.0	0.50
8	48.0	0.43	23	25.0	-0.03	38	58.0	0.56
9	48.0	0.56	24	55.0	0.63	39	51.0	0.23
10	66.0	0.60	25	80.0	0.20	40	88.0	0.10
11	73.0	0.46	26	68.0	0.43	41	31.0	0.43
12	66.0	0.40	27	55.0	0.36	42	66.0	0.40
13	23.0	0.33	28	50.0	0.46	43	51.0	0.43
14	70.0	0.13	29	46.0	0.20	44	18.0	0.30
15	48.0	0.30	30	60.0	0.53	45	56.0	0.53

Difficulty Index= DIFI
Discrimination Index= DI

Results are presented in Table 1. Scores of 89 students were analyzed. The mean score was 34.93±3.72, ranging from 12 to 41. The mean number of students who achieved higher scores on the MCQs test was 18.17±2.94 while the mean number of students who scored low on test was 26.56±7.50. Mean DIFI was 58.66±16.71, whereas the mean Discrimination Index (DI) was 0.37±0.17.

Each MCQ was analyzed for DIFI, and DI was calculated and presented in Table 2.

Negative discrimination values were obtained for Question No. 3 and 23. Question 40 has a maximum of 88% DIFI, and question no 13 has a minimum DIFI of 23%. DIFI categorization of MCQs is presented in Figure 1.

Of 45 MCQS, 68.9% had a good/acceptable level of difficulty and were stored, whereas 24.4% were too easy and 6.7% were too complex and confusing, and they were placed for revision.

Figure 2 describe the assessment of MCQs based on DI. Out of 45 MCQS, 27 (60%) were found to have excellent discrimination and were stored whereas 13 (29%) were having good discrimination and were reviewed and stored for reuse. While 5 (11%) MCQs having poor discrimination were reviewed and after modifications 3 were stored and 2 were discarded.

Results of DE was 93.32±19.60 presented in Table 3. Out of 135 distractors, 129 were functional while NFD were 6.

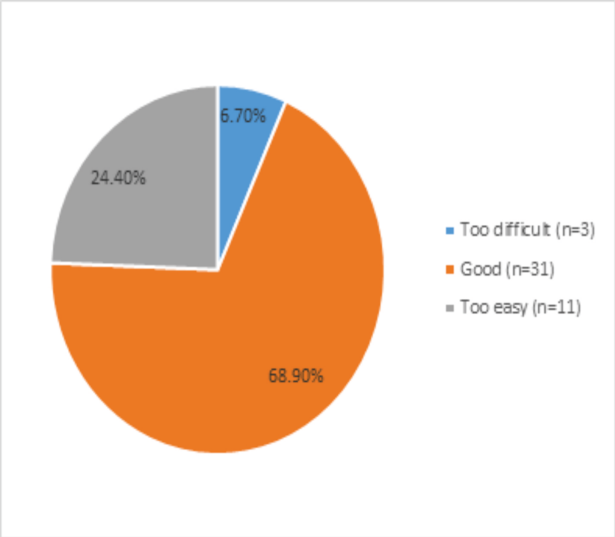

Fig 1: Difficulty index of items

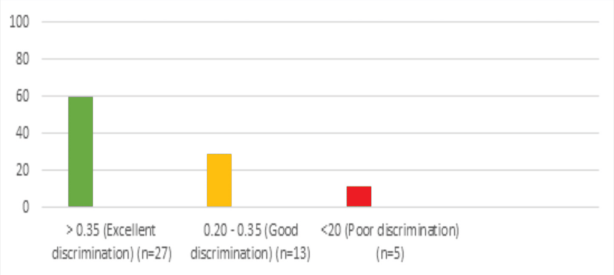

Fig 2: MCQ assessment based on DI in percentage

Table 3: Distractor analysis of MCQ items (n = 45)

Item description	n	%
Total Distractors	135	100
Functional Distractors	129	95
Non-Functional Distractors	6	5
Items with DE = 100 % (0 NFD)	39	86.7
Items with DE = 66.6 % (1 NFD)	4	8.9
Items with DE = 33.3 % (2 NFDs)	1	2.2
Items with DE = 0 % (3 NFDs)	1	2.2
Distractor Efficiency (Mean ± SD)	93.32±19.60	

DE= Distractor Efficiency
NFD= Non-functioning distractor

Discussion

Quality of assessment is of prime importance in competence-based education. A valid and reliable assessment improves learning, thereby ensuring competence.[12] Commonly used method for the assessment of cognitive competence of students in medical curriculum is carried out by MCQs. The assessment needs to be valid and reliable for exam effectiveness. It is, therefore essential to construct quality MCQs for an effective and trustworthy assessment. Quality of MCQs is quite easily performed by item analysis and is generally assessed

by three parameters: DIFI, DI, and DE.[13]

In our study of DIFI, 68.9% were in the acceptable range of 30-70%, while 24.4% and 6.7% were too easy and too hard, respectively. DI was 0.37±0.17. DE was 93.32±19.60, demonstrating 86.7% functional distractors and 13.3% NFDs. Similar results were demonstrated in study of Kaur M et al., where DIFI analysis found 76% of items in the acceptable range, 22% were too easy, and 2% of item were too difficult. Results of DI found 62% of items as excellent, 24% of items as good, and 14% of items as poor. DE analysis demonstrated 82% functional distractors and 18% NFDs.[13] Results of a study by Rao C et al. showed 85% of MCQs in the desirable range while 5% and 10% were found easy and difficult, respectively.[11] DI is a measure of skilled and unskilled students.[16] DI value varies between 0 and 1. DI value of <0 may also be obtained when it is termed as negative DI.[19] This happens when a higher number of students of the lower achiever group mark the item correctly than of high achiever group. The reason for a negative DI is due to an ambiguous question or wrong answer key. In our study, out of 45 MCQS, 27 were found to have excellent discrimination and were stored. Good discrimination items were 13 and were therefore kept for revision and to be stored. While 5 MCQs were found to have poor discrimination and were kept for revision and possible storage. Question no. 3 and 23 having negative discrimination were discarded. Our study results of DI are comparable to a study by Patel and Mahajan, Patel and Patel, Singh et al. and Mehta and Mokhasi et al. showing 18% MCQs with DI < 0.2 while 82% MCQs had DI value of > 0.20; 76% good to excellent with 24% having poor DI; 30% item with DI < 0.2 and 70% with DI >0.2; and mean DI was 0.33 ± 0.18 and out of total 50 items, 15 items had DI < 0.2 and 35 items had DI > 0.20 respectively.[17,20,21] Mean DE of our study was 93.32±19.60 comparable to the study by Gajjar et al. with an overall mean DE of 88.6 ± 18.6; the study by Patil and Patil with a mean DE of 82.8 ± 15.6.[17] Results of a study in Pakistan by Beenish et al., are also comparable to our study demonstrating mean DE of 85.33%±21.69%.[22] The results of different studies reflect the presence of many NFDs in the MCQ examination.[12] Developing an effective distractor is the most challenging part in quality MCQ construction.[23] Ideal one best type MCQ should have average difficulty, high discrimination, and three functional distractors.[24] It is presumed that increasing or decreasing the number of distractors have an impact on quality of MCQ, being improved with an increased number thereby reducing the chance of guessing while decreasing the number improves guessing. Research in this regard shows no difference in the quality of MCQs with four or five items.[12,25] It is opined that too difficult or too easy MCQs may result in unreal marking that may become a source of decreased motivation and should be discarded.[17] However, keeping in view that construction of MCQ being time-consuming and laborious job, MCQs in too difficult and too easy range in our study were kept for revision and subsequent storing after reconstruction.

Recommendation

Faculty development programs (FDPs) improve the faculty skills in item construction.[12] This is stressed that a single FDP is of limited value in improving the quality of MCQs.[26] It is therefore appropriate to have repeated FDPs for desirable improvement in item writing skills.

Limitations

A small number of MCQs were analyzed. Further research with a large number of items is recommended for better and improved validity and reliability of test items.

Conclusion

MCQ item analysis is a valued simple tool for identifying the deficiencies and flaws of MCQ for their reconstruction and subsequent reuse. Improvement in the quality of MCQs improves the reliability of tests. Item analysis also improves the skills of item construction.

REFERENCES

1. Perez G, Kattan E, Collins L, Wright AC, Rybertt T, Gonzalez A, et al. Assessment for learning: experience in an undergraduate medical theoretical course. Revista Medica de Chile. 2015; 143: 329-36. doi: 10.4067/S0034-98872015000300007
2. Retrieved 02-10-2021 from https://www.queensu.ca/teachingandlearning/modules/assessments/04_s1_01_intro_section.html
3. Peeters MJ. Targeting assessment for learning within

pharmacy education. American Journal of Pharmaceutical Education. 2017; 81: 6243. doi: 10.5688/ajpe6243

4. Schut S, Heeneman S, Bierer B, Driessen E, van Tartwijk J, van Der Vleuten C. Between trust and control: Teachers' assessment conceptualisations within programmatic assessment. Medical Education. 2020; 54: 528-37. doi: 10.1111/medu.14075
5. Retrieved 03-06-2022 from https://cft.vanderbilt.edu/guides-sub-pages/blooms-taxonomy/. Armstrong, P. (2010) Bloom's Taxonomy. Vanderbilt University Center for Teaching.
6. 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 University Medical Journal. 2018; 18: e68–e74. doi: 10.18295/squmj.2018.18.01.011
7. Kumar D, Jaipurkar R, Shekhar A, Sikri G, Srinivas V. Item analysis of multiple choice questions: A quality assurance test for an assessment tool. Medical Journal Armed Forces India. 2021; 77: S85-9. doi: 10.1016/j.mjafi.2020.11.007
8. Song Y, Kim H, Park SY. An item response theory analysis of the Korean version of the CRAFFT scale for alcohol use among adolescents in Korea. Asian Nursing Research. 2019; 13: 249-56. doi: 10.1016/j.anr.2019.09.003
9. Tavakol M, Dennick R. Postexamination analysis: The item characteristic curve. Academic Medicine. 2018; 93: 811. doi: 10.1097/ACM.0000000000002110
10. Abdulghani HM, Ahmad F, Irshad M, Khalil MS, Al-Shaikh GK, Syed S, et al. Faculty development programs improve the quality of Multiple Choice Questions items writing. Scientific reports. 2015; 5: 9556. doi: 10.1038/srep09556
11. 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; 2: 201-4. doi: 10.4103/2395-2296.189670
12. Belay LM, Sendekie TY, Eyowas FA. Quality of multiple-choice questions in medical internship qualification examination determined by item response theory at Debre Tabor University, Ethiopia. BMC Medical Education. 2022; 22: 1-1. doi: 10.1186/s12909-022-03687-y
13. Kaur M, Singla S, Mahajan R. Item analysis of in use multiple choice questions in pharmacology. International Journal of Applied and Basic Medical Research. 2016; 6: 170-3. doi: 10.4103/2229-516X.186965
14. Mehta G, Mokhasi V. Item analysis of multiple choice questions – An assessment of the assessment tool. International Journal of Health Sciences and Research. 2014; 4: 197–202.
15. Kowash M, Hussein I, Halabi MA. Evaluating the Quality of Multiple Choice Question in Paediatric Dentistry Postgraduate Examinations. Sultan Qaboos University Medical Journal 2019; 19: e135–e41. doi: 10.18295/squmj.2019.19.02.009
16. Bhat SK, Prasad KH. Item analysis and optimizing multiple-choice questions for a viable question bank in ophthalmology: A cross-sectional study. Indian journal of ophthalmology. 2021; 69: 343-6. doi: 10.4103/ijo.IJO_1610_20.
17. Patil VC, Patil HV. Item analysis of medicine multiple choice questions (MCQs) for under graduate (3rd year MBBS) students. Research Journal of Pharmaceutical Biological and Chemical Sciences. 2015; 6: 1242-51.
18. Sajjad M, Iltaf S, Khan RA. Nonfunctional distractor analysis: An indicator for quality of Multiple choice questions. Pakistan Journal of Medical Sciences. 2020; 36: 982-6. doi: 10.12669/pjms.36.5.2439.
19. 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. 2014; 39: 17–20. doi: 10.4103/0970-0218.126347
20. Patel KA, Mahajan NR. Itemized analysis of questions of multiple choice question exam. International Journal of Science and Research. 2013; 2: 279-80. doi: 10.15373/22778179/FEB2013/94
21. Singh JP, Kariwal P, Gupta SB, Shrotriya VP. Improving Multiple Choice Questions (MCQs) through item analysis: An assessment of the assessment tool. International Journal of Sciences & Applied Research. 2014; 1: 53-7.
22. Mehmood B, Sultan S, Azhar M, Shakoor A, Khan NB, Niazi Z. Quality analysis of multiple choice questions. Pakistan Oral & Dental Journal. 2020; 40: 236-9.
23. Touissi Y, Hjiej G, Hajjioui A, Ibrahimi A, Fourtassi M. Does developing multiple-choice questions improve medical students' learning? A systematic review. Medical Education Online. 2022; 27: 2005505. doi: 10.1080/10872981.2021.2005505
24. Hingorjo MR, Jaleel F. Analysis of one-best MCQs: the difficulty index, discrimination index and distractor efficiency. JPMA-Journal of the Pakistan Medical Association. 2012; 62: 142-7.
25. Fozzard N, Pearson A, du Toit E, Naug H, Wen W, Peak IR. Analysis of MCQ and distractor use in a large first year

Health Faculty Foundation Program: assessing the effects of changing from five to four options. BMC medical education. 2018; 18: 1-0. doi: 10.1186/s12909-018-1346-4

26. Gupta P, Meena P, Khan AM, Malhotra RK, Singh T. Effect of faculty training on quality of multiple-choice questions. International Journal of Applied and Basic Medical Research. 2020; 10: 210-14. doi: 10.4103/ijabmr.IJABMR_30_20