ORIGINAL ARTICLE

Difficulty Index, Discrimination Index and Distractor Efficiency in Multiple Choice Questions

Wajiha Mahjabeen[1], Saeed Alam[2], Usman Hassan[3], Tahira Zafar[4], Rubab Butt[5], Sadaf Konain[6], Myedah Rizvi[7]

Author`s Affiliation

[1]Associate Professor-Pathology
[2]Professor of Pathology
[3-7]Demonstrator-Pathology
Islamabad Medical and Dental College, Islamabad

Author`s Contribution

[1]Conception, synthesis, planning of research, manuscript writing and Data Analysis
[2]Conception, Review of study
[3-7]Collection of data, Data Analysis

Address of Correspondence
Dr. Wajiha Mahjabeen
doctor_wajeeha@yahoo.com

ABSTRACT

Objective: Our first objective was to evaluate the quality of MCQs by analyzing difficulty index, discrimination index and distractor efficiency. Our second objective was to find out the association of MCQs having good difficulty and discrimination indices with distractor efficiency.

Methodology: This cross-sectional study was conducted at department of Pathology, Islamabad medical and dental college. Midterm paper comprising of total 65 MCQs was assessed for difficulty index (DIF), discrimination index (DI) and distractor efficiency (DE). Data was entered in Microsoft Excel 2010 and SPSS 21. Quantitative variables were expressed as mean±SD. Qualitative variables were written as number and percentage. Independent t-test was applied to find out the association of DIF and DI with DE.

Results: According to DIF, out of total 65, 53(81%) MCQs were in acceptable category, only 1(2%) MCQ was too difficult and 11(17%) were too easy. Regarding DI, total 34(62%) MCQs showed excellent discrimination tendency to distinguish low and high performer students. While 15(23%), 5(8%) and 11(17%) MCQs demonstrated good, acceptable and poor discrimination ability respectively. Out of total 260 distractors, 72% were functional and only 28% were non-functional. Total 16(25%) MCQs had zero non-functional distractor (NFDs), while 30(46%) and 16(25%) MCQs had 1 and 2 NFDs respectively. Only 3(5%) MCQs were with 3 or more NFDs. DE was significantly more (100%) in 1 difficult item as compared to 11 easy items in which DE was less (36.33%). However, DE in MCQs having poor and good DI was almost same.

Conclusion: In this paper of Pathology, large number of MCQs have acceptable level of DIF (81%) and DI (83%). Distractor efficiency related to presence of zero or 1 NFD is 71%. Through item analysis, standardized MCQs having average DIF, high discrimination power with large number of functioning distractors can be developed. Thus, it is an effective way to improve the validity of examination and to efficiently assess the student performance.

Key words: Difficulty index, discrimination index, distractor efficiency, multiple choice questions

Introduction

In different professional examinations use of multiple choice questions is frequently increasing to assess the knowledge of students.[1] Well-constructed MCQ is a useful examination tool that can cover the wide area of subject with objectivity across all cognitive levels.[2] It also lessens the evaluator's bias by minimizing individual's judgement during scoring. Development of standardized MCQ is a time-consuming task. If a MCQ is not well constructed, it

can be easier or more difficult to be attempted by students as required. If the options given in MCQ are not according to standardized criteria, it will reduce the student recalling, comprehension or problem-solving skills and will direct the students towards guessing. [3-5] In medical colleges it is very important to give adequate and accurate knowledge to students and to improve their practical skills. A medical student should be more inquisitive and more analytical to develop appropriate professional attitude. The purpose of assessment taken during teaching and learning practice is multifold. It not only assure the students capability to grasp the knowledge given but also to observe that how much our teaching strategies are effective. Therefore process of assessment should be effective and trustworthy.[6] In order to improve the students' knowledge and to enhance the quality of examination, continuous analyses of student's assessment methodologies should be a key step. There are previously defined pre-validation and post validation assessment methods to analyze the formulated questions. In the process of pre-validation, before conduction of assessment a group of subject specialists should evaluate the applicability of topics covered in paper and appropriateness of structure of MCQs including stem and options. Post validation process is basically a statistical method that is also called as item analysis. This is a valuable, relatively simple but an effective process to check the reliability and validity of MCQs.[7, 8] This is helpful in three aspects. First of all it tells that MCQ given to student is difficult or easy to attempt that is called the difficulty index (DIF). Secondly it can discriminate the students having good knowledge about subject assessed from those not performing well. It is called as discrimination index (DI). Thirdly it helps the subject specialist to assess the credibility of incorrect options (distractors). This is known as distractor efficiency (DE). Overall this analysis gives guidelines to evaluator to amend the MCQs before next examination to make it more appropriate.[9, 10]

In our setup, mostly medical teachers are not able to assess the quality of their MCQs through item analysis. As a result, many unstandardized MCQs can be added in examinations. We performed this study to evaluate the quality of MCQs by analyzing DIF, DI and DE. Our second objective was to find out the association of MCQs having good difficulty and discrimination indices with DE. This will be helpful to assess how non-functioning distractors affect the ideal questions.

Methodology

This cross-sectional study was conducted in the department of pathology at Islamabad Medical and Dental College Islamabad during the academic session 2017. Total 110 students of 4th-year MBBS appeared in Pathology midterm examination. Before assessment paper was evaluated by a subject specialist. Paper was comprised of 65 MCQs, each having a single stem with five options including one correct answer and four distractors (incorrect answers). Each MCQ was assigned one mark. Maximum marks possible to score were 65 and minimum was zero, with no negative marking. For item analysis, results of all papers were ranked in descending order, from highest marks to lowest marks. Then papers were divided into quartiles. Upper quartile or high scored (n=33) and lower quartile or low scored (n=32) groups were included into the analysis. Paper with average scores, middle quartiles (n=45) were excluded from the study. DIF, DI and DE were calculated to evaluate the MCQs.

DIF represents the percentage of students who correctly answer the questions. A higher value of DIF shows that increased number of students gave the correct answer. It indirectly proves that questions are easy to attempt. The range of DIF is from 0-100%. Following formula is used to calculate the DIF

$$DIF = [(H+L)/N] \times 100$$

H= Number of students gave correct options in high score group

L=Number of students gave correct options in low score group

T=Total number of students in both groups

Criteria of categorization in DIF is: DIF>70%=Too easy,

DIF b/w 30-70%=Average,

DIF b/w 50-60%= Good,

DIF<30%=Too difficult

DI is the capacity of a MCQ to differentiate the students getting high scores from low performing ones. Its range is 0-1. Formula used to calculate DI is

$$DI = 2 \times [(H-L)/N] \times 100$$

DI is categorized as:

DI≤0.2= Poor,

DI b/w 0.21-0.24= Acceptable,

DI b/w 0.25-0.35= Good,

DI≥0.36=Excellent

DE is the ability of incorrect answers to distract the students. If < 5% students choose the incorrect answers, it is called non-functioning distractor (NFD). Distractors selected by >5% of students is called functional distractors (FD). The range of DE is 0-100%.

DE is categorized on the basis of the number of NFD present in a MCQ. If MCQ has 3 or more NFDs, its DE is 0%. DE is labeled as 33.3%, 66.6% and 100% on the basis of the presence of 2, 1 or none NFD in an MCQ. [11, 12]

Data was entered in Microsoft Excel 2010 and SPSS 21. Quantitative variables were expressed as Mean±SD. Qualitative variables were written as number and percentage. Independent t-test was applied to find out the association of DIF and DI with DE.

Results

Out of 110 students, total 65 were categorized as high performers and low performers. Papers of these 65 students were included for analysis. Each paper was comprised of total 65 MCQs and 260 distractors. According to mean, both DIF and DI of total MCQs were in good category. Mean±SD and range of DIF, DI and DE have been shown in table I.

Table I: Characteristics of MCQs evaluation criteria's	
Parameters	Result
Students (n)	65
MCQs (n)	65
Score Total (n)	65
Score obtained (Mean±SD) Range	38.18±12.17 12-62
Difficulty index (%) Mean±SD Range	58.74±14.39 12.31-86.15
Discrimination index Mean±SD Range	0.35±0.16 .03-.68
Distractor efficiency (%) Mean±SD Range	63.55±27.47 0-100

Regarding DIF, out of total 65, Majority of MCQs were in the acceptable category (Figure 1). Among these acceptable category MCQs (n=53), 18 fall under the category of having good DIF. A large number of MCQs showed excellent (52%) and good (23%) discrimination tendency to distinguish low and high performer students (Figure 2).
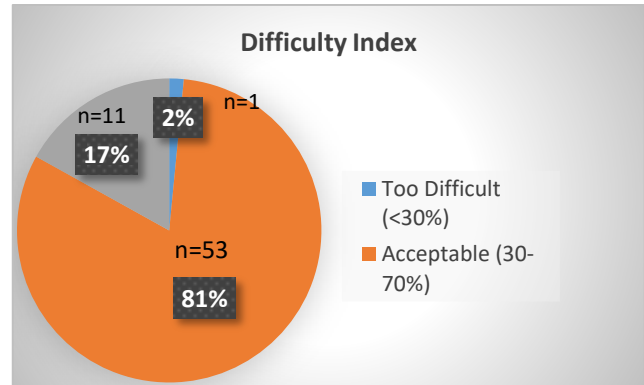

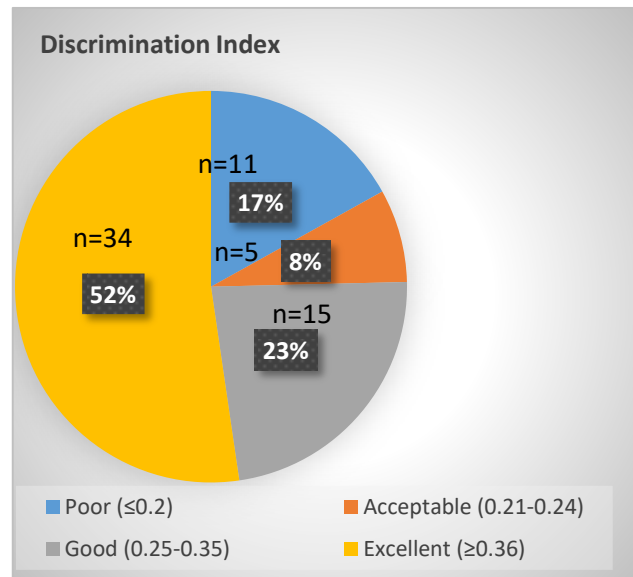**Figure 1: Categorization of MCQs according to difficulty index (n=65)**

Figure 2: Categorization of MCQs according to discrimination index (n=65)

Out of total 260 distractors, 72% were functional and only 28% were non-functional. Total 71% (n=46) MCQs showed DE up to 66.6%. (Table II). DE was significantly more (100%) in 1 difficult item as compared to 11 easy items in which DE was less (36.33%). However, DE in MCQs having poor and good DI was almost same (Table III).

Table II: Number of distractors and categorization of MCQs according to distractor efficiency

Parameters	Number (%)
MCQs (Total)	65
Distractors (Total)	260
Functional Distractors	188 (72)
Non-Functional Distractors	72 (28)
MCQs with zero NFDs/ 4 FDs (DE=100%)	16 (25)
MCQs with 1 NFDs / 3 FDs (DE=66.6%)	30 (46)
MCQs with 2 NFD s / 2 FDs (DE=33.3%)	16 (25)
MCQs with 3 or more NFDs / 1 or 0 FDs (DE=0%)	3 (5)

were 3(10%) MCQs, which overall fulfilled the standards of ideal MCQs.[15]

In our study, mean and standard deviation for DIF, DI and DE were fallen in the category of good MCQs. These results are almost similar to a study analyzed 30 MCQs showing mean DIF and DE in the range of good MCQs. Their DI fall in the category of excellent MCQ.[16] Another study done item analysis of 40 MCQs, revealed mean DIF and DI almost in accordance to our study (average MCQs) with high mean DE (excellent MCQs).[17] In a study conducted at Ghana analysis of 50 test items revealed that mean DIF and DE was average and DI was an acceptable group.[18]

Regarding difficulty index, in our study out of total 65 MCQs, 1(1%) MCQ was too difficult and 11(17%) were

Table III: Association of distractor efficiency with difficulty index and discrimination index

	Difficulty index			Discrimination index		
	Difficult (<30%)	Easy (>70%)	p-value	Poor (≤0.2)	Good & Excellent (≥0.25)	p-value
Number of MCQs	1	11		11	49	
Distractor efficiency % (Mean±SD)	100	36.33±23.33	0.026	60.58±38.93	63.90±23.42	0.711

Discussion

MCQ with single correct answer is an effective way to assess the student's cognitive knowledge. According to blooms taxonomy, a well-constructed MCQ is an efficient tool to quickly evaluate different levels of cognition like comprehension, application, analysis, and synthesis among students.[13] However, the first mandatory step for quality assessment is standardization of MCQs. Frequent evaluation of questions through item and test analysis is an active approach to make the valid pool of MCQs.

For an ideal MCQ level of difficulty should be average with (30-70%) with high DI (>0.25) and 100% DE.[11, 14] In our study according to DIF criteria, out of total 65 MCQs, the majority (81%) fulfills the criteria of an ideal MCQ. As per the DI and DE, total 52% and 25% fall in the categorization of ideal MCQ. There were total (10)15% MCQs which satisfied all the three criteria's of ideal MCQs. Our results are comparable to a study conducted at India. In that study out of 30 MCQs 15(50%) fulfilled the criteria of DIF for ideal MCQs, while 21(70%) and 17(57%) MCQs satisfied the criteria of DI and DE for an ideal MCQ. There

too easy. Total 53 (81%) and 18(28%) MCQs were in acceptable and good category respectively. Results are comparable to another study analyzing 40 MCQs. Study revealed that 7(17.5%) MCQs were too easy and 3(7%) were too difficult. Remaining 18(45%) and 12(30%) were fall in acceptable and ideal category respectively.[17] Another study conducted in 2016 showed that out of total 30 MCQs, 5(17%) were very easy and 11(37%) were too difficult. Remaining 4(13%) and 10(33%) fall in the category of good and very good respectively.[15] A study conducted at India in 2017 accessed 5 papers comprising of total 200 MCQs. The study revealed that 74(37%) MCQs were too difficult and 33(16%) were too easy. Remaining 93(46%) MCQs were in the average category.[19] Another study analyzing MCQ paper consisting of 30 questions showed that 2(7%), 24(80%) and 4(13%) MCQs were too easy, acceptable and too difficult respectively.[16]

In the present study, concerning discrimination tendency, 34(52%) MCQs showed excellent predisposition to distinguish students gaining low and high marks. While 15(23%), 5(8%) and 11(17%) MCQs demonstrated good,

acceptable and poor discrimination ability respectively. A study conducted in 2017 analyzed discrimination tendency of 40 MCQs. The study demonstrated that 17(42.5%) and 7(17.5%) MCQs had excellent and good discrimination tendency respectively. While 1(2.5%) MCQ fall in acceptable range and 15(37.5%) had poor tendency to discriminate the low and high performers.[17] In a medical college at India, analysis of 30 MCQs revealed that 9(30%) had poor discrimination tendency. While 6(20%) and 15(50%) MCQs were categorized as having good and very good tendency to discriminate students on performance basis.[15] A study analyzing discrimination tendency of questions given in 5 tests showed that out of 200 MCQs, 79 (39.5%) had poor while 47(23.5%), 13(6.5%) and 61(30.5%) MCQs had marginal, good and excellent DI respectively.[19] A study conducted at Govt medical college in India categorized 30 MCQs into 8(27%), 3(10%) and 19(63%) as having poor, good and excellent DI.[16]

The present study showed that in 16 (25%) MCQs all four wrong options fully distracted the student's attention. While 30(46%) and 16(25%) MCQs had 3 and 2 functional distractors respectively. Only 3(5%) MCQs were with one functional distractor. Results are in accordance with study conducted in 2017 showing 8(26.6%) MCQs with all three functioning distractors. Total 13(43.33%), 7(23.33%) and 2(6.66%) items had three, two and zero functioning distractors respectively.[16] A recent study analyzed 30 MCQ with 90 distractors showed that 17(56.7%) items were with zero NFD (three functional distractors) and remaining 3 and 10 were with 2 NFD (two functional distractor) and 1NFD (two functional distractors) respectively.[15] Similarly another recent study analyzing 40 MCQ with total 120 distractors, revealed that a number of items with three functional (0 NFD, DE=100%) distractors were high (26 [65%]). While items with two functional (1 NFD, DE=33%) distractors and with one functional (2 NFDs, DE=66%) distractor were 10(25%) and 4(10%) respectively.[17]

Conclusion

In this paper of Pathology, large number of MCQs have acceptable level of DIF (81%) and DI (83%). Distractor efficiency related to presence of zero or 1 NFD is 71%. Development of standardized MCQs having average DIF,

high discrimination power with large number of functioning distractors is an effective way to improve the validity of examination. It can also efficiently assess the student performance. For this quality assessment process, conduction of faculty development program can be helpful to enhance the learning and performance of medical faculty for development of new standardized MCQs.[20, 21]

References

1. Hubbard JP, Clemans WV. Multiple-choice examinations in medicine. A guide for examiner and examinee. Academic Medicine. 1961;36(7):848.
2. Loh KY, Elsayed I, Nurjahan M, Roland G. Item difficulty and discrimination index in single best answer MCQ: end of semester examinations at Taylor's clinical school. Redesigning learning for greater social impact: Springer; 2018. p. 167-71.
3. Haladyna T. Selected-response format: developing multiple-choice items. 1st ed. chapter 5; New York, NY, Routledge, 2013; 79-131
4. Kuechler WL, Simkin MG. Why is performance on multiple-choice tests and constructed-response tests not more closely related? Theory and an empirical test. Decision Sciences Journal of Innovative Education. 2010;8(1):55-73.
5. Palmer EJ, Devitt PG. Assessment of higher order cognitive skills in undergraduate education: modified essay or multiple choice questions? Research paper. BMC Medical Education. 2007;7(1):49.
6. Tavakol M, Dennick R. Post-examination analysis of objective tests. Medical Teacher. 2011;33(6):447-58.
7. Collins J. Education techniques for lifelong learning: writing multiple-choice questions for continuing medical education activities and self-assessment modules. Radiographics: a review publication of the Radiological Society of North America, Inc. 2006;26(2):543-51.
8. Considine J, Botti M, Thomas S. Design, format, validity and reliability of multiple choice questions for use in nursing research and education. Collegian. 2005;12(1):19-24.
9. Tarrant M, Ware J, Mohammed AM. An assessment of functioning and non-functioning distractors in multiple-choice questions: a descriptive analysis. BMC Medical Education. 2009;9(1):40.
10. Rasiah S-MS, Isaiah R. Relationship between item difficulty and discrimination indices in true/false-type multiple choice questions of a para-clinical multidisciplinary paper. Annals Academy of Medicine Singapore. 2006;35(2):67-71.
11. Ananthakrishnan N. Item analysis-validation and banking of MCQs. N Ananthkrishnan, KR Sethuraman & S Kumar Medical Education Principles and Practice Pondichery: JIPMER. 2000:131-7.
12. Chauhan PR, Ratrhod SP, Chauhan BR, Chauhan GR, Adhvaryu A, Chauhan AP. Study of Difficulty Level and Discriminating Index of Stem Type Multiple Choice Questions of Anatomy in Rajkot. Biomirror. 2013;4(6).
13. Carneson J, Delpierre G, Masters K. Designing and managing MCQs. Appendix C: MCQs and Blooms taxonomy (Online) 2011 (Cited 2011 Feb2). Available from URLhttp. web uct ac za/projects/cbe/mcqman/mcqappc html.

14. 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(2):142.
15. Patil R, Palve SB, Vell K, Boratne AV. Evaluation of multiple choice questions by item analysis in a medical college at Pondicherry, India. International Journal Of Community Medicine And Public Health. 2017;3(6):1612-6.
16. Ingale AS, Giri PA, Doibale MK. Study on item and test analysis of multiple choice questions amongst undergraduate medical students. International Journal Of Community Medicine And Public Health. 2017;4(5):1562-5.
17. Patel RM. Use of Item analysis to improve quality of Multiple Choice Questions in II MBBS. Journal of Education Technology in Health Sciences. 2017;4(1):22-9.
18. Quaigrain K, Arhin AK. Using reliability and item analysis to evaluate a teacher-developed test in educational measurement and evaluation. Cogent Education. 2017;4(1):1301013.
19. 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. International Journal Of Community Medicine And Public Health. 2017;4(6):1876-81.
20. Tenzin K, Dorji T, Tenzin T. Construction of multiple choice questions before and after an educational intervention. Journal of the Nepal Medical Association. 2017;56(205).
21. 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.