JBACHS
Journal of Basic and Clinical Health Sciences

J Basic Clin Health Sci 2024; 8: 490-497
https://doi.org/10.30621/jbachs.1427464
RESEARCH ARTICLE

ITEM ANALYSIS IN MULTIPLE CHOICE QUESTIONS: A STUDY ON QUESTION DIFFICULTY AND AUTHORS' EVALUATION

Serap Konakci[1]

[1]Dokuz Eylul University, Faculty of Medicine, Department of Medical Education, Izmir-Turkey

ORCID: S.K. 0000-0002-3325-6382

Corresponding author: Serap Konakcı, **E-mail:** serapkonakci@yahoo.com

Received: 29.01.2024; **Accepted:** 10.05.2024; **Available Online Date:** 31.05.2024

Cite this article as: Konakci S. Item Analysis in Multiple Choice Questions: A Study on Question Difficulty and Authors' Evaluation. J Basic Clin Health Sci 2024; 8: 490-497.

ABSTRACT

Introduction: Multiple choice questions (MCQs) are widely used in medical education. This study aims to evaluate the quality of MCQ and the predictions of MCQ authors about the difficulty levels of their questions.

Methods: In this study, the Difficulty Index (DIF I), Discrimination Index (DI), and Distractor Efficiency (DE) values of 688 MCQs in the exams held in the first year of the 2021-2022 academic year of Dokuz Eylül University Faculty of Medicine were investigated. The data were reported as a percentage and mean ± standard deviation (SD), minimum and maximum values of items. DIF I and DI values among the groups formed according to the distractor activity were compared using the t-test, and the effect size was calculated. Estimated and actual DIF I, one-to-one matching was evaluated with Mc neamer chi-square test.

Results: The results of our research are based on the analysis of data from 688 MCQs. DIF I mean was 0.57±0.21, and 47.5% was at the ideal difficulty level. There was a significant difference between the estimated and the actual DIF I ($p=0.002$). The DI average was 0.31 ± 0.17, and the discrimination level of 43.6% was excellent. 36.8% of distractors were NFD. MCQ's difficulty and discriminatory ability were significantly different according to the number of NFDs ($p<0.001$ for all).

Discussion: It was determined that the number of NFDs significantly affected difficulty and discriminatory ability. There was a difference between the estimates of the difficulty and the actual values. Reducing the number of options in MCQs and being more careful when crafting questions can improve the quality of the questions.

Keywords: Difficulty index, Discrimination index, Functioning distractors, Item Analysis, MCQs.

INTRODUCTION

Tests consisting of multiple choice questions (MCQ) are widely used in medical education, as they allow the evaluation of high-level cognitive areas of Bloom's taxonomy and the evaluation of a large number of people simultaneously (1-5).

Item analysis allows the quality of MCQs to be evaluated. Difficulty Index (DIF I), Discrimination Index (DI), and Distractor efficiency (DE) are the most frequently used item analysis values. With these analyses, the properties of each substance can be determined separately (6). Also, item analysis helps to decide on the selection, revision, or removal of questions to create question banks. It provides data for question writers on their performance and guides them to write more effective MCQs. Guides and studies on MCQ preparation can be easily found in the literature (1, 7-12).

DIF I takes a value between '0' and '1'; the closer it is to 1, the easier the MCQ is. DIF I value for medical education can be grouped as ≤ 0.29 too difficult, 0.30-0.70 acceptable, 0.50-0.60 ideal, and ≥ 0.70 too easy (13-15).

DI defines the extent to which the item can distinguish between students knowledgeable in the targeted field and students who are not. DI can take a value between '-1' and '+1'. As DI approaches '+1', the ability to distinguish between those who know and those who do not know increases. If the DI value of an item is 0.19 or below and does not contain an obvious error that can be corrected when examined, it is recommended that the item be removed from the test/not used again. A DI value of ≥0.35 is considered excellent in an ideal test.

DE is effective in determining the difficulty and discrimination level of an item. For someone who does not have sufficient knowledge of the subject being evaluated, distractors are expected to be the correct answer and be preferred. Arranging the appropriate distractor is as tricky as arranging the correct response (16). A distractor with a preference rate of <5% is generally considered non-functional. The more non-functional distractors (NFD) in an MCQ, the lower and easier it becomes to discriminate.

It is important in ensuring the validity and reliability of the exams that MCQ writers master the principles of question preparation and have knowledge about interpreting the results of item analysis. Besides, it is important to consider the item analysis results in creating question banks.

The questions of this research are as follows:

- What are the conditions of DIF I, DI, and DE of the evaluated MCQs?
- Does the number of NFD have an effect on DIF I and DI?
- Do question authors have a realistic foresight about the difficulty level of their questions?

In this study, the evaluation of MCQ quality and the assessment of question authors' predictions about the difficulty level of their questions are aimed to find answers to the research questions mentioned above.

MATERIALS AND METHODS

In the 2021-2022 academic year, a total of six MCQ tests were applied to 346 students studying at Dokuz Eylül University Faculty of Medicine (DEUFM) Term 1 for knowledge evaluation throughout the year. The tests were prepared using the blueprint. The number of questions to be included in the tests varies depending on the block time and total number of targets (min: 100 - max: 125 MCQs). MCQs in the tests have five options and one correct answer. All of the questions were used for the first time. Item analysis was performed routinely after each exam. The results were used as a guide to decide on the reuse of questions stored in the question bank. In the 2021-2022 academic year, MCQ authors recorded their estimation of DIF I as 'very easy, acceptable or very difficult' when preparing the question. Before all MCQs were used in the tests, they were reviewed by an evaluator other than the MCQ authors in terms of grammatical clues, logical clues, having more details in the right option, the arrangement of options (chronological or numerical order), and unnecessary information in the stem. After the necessary arrangements and corrections were made, they were used in the exams.

In our faculty, a question discussion session was held after each exam. Questions and correct answers were shared with students in discussion sessions. After these sessions, students had the right to object to the information contained in the MCQs and the correct answer by citing literature. According to the item analysis results, questions with a known rate of 10% or less, questions with a noticeably high rate of marking a particular distractor, questions with DI ≤0.19 and DIF I value <30, and questions objected to by students citing literature support were consulted with the MCQ author.

A total of 700 MCQs were used in six Term I tests that took place during the period covered by the research, and the authors of 40 MCQs were consulted in line with the criteria listed above. Twenty of the MCQs whose authors were consulted were excluded from the evaluation because they were found to contain informational errors. This process was routinely applied in all MCQ exams in our faculty, and the final calculation of student scores was made after these procedures. The research results were based on the evaluation of data from the item analysis of 688 MCQs used in calculating student scores.

Item Analysis

DIF I and DI are calculated and categorized as follows:

H= Number of students giving correct responses in the high score group (upper 27%).
L= Number of students giving correct responses in the low score group (lower 27%).
N= Total no of responses in both groups.

DIF I = [(H+L) / N] (DIF I of an item range between 0-1)
Criteria for categorization of DIF I is,

DIF I ≥0.7 = too easy
DIF I = 0.3 – 0.7= acceptable
DIF I = 0.5-0.6= ideal
DIF I ≤0.29= too difficult

DI = 2 x [(HL) / N] [DI of an item range between (-1) – (+1)]
Criteria for categorization of DI are,

DI ≤ 0.2 = poor
DI = 0.21-0.24 =acceptable
DI = 0.25-0.34=good
DI ≥0.35 = excellent

For DE, a distractor with a preference rate of <5% was considered non-functional, and the number of non-functional distractors (NFD) was determined for each MCQ.

Statistical Analysis

Statistical analyses of the research were carried out via SPSS v.24.0 (IBM, Armonk, NY, United States of America) using the item analysis results of the education management system used in our faculty. The data were reported as a percentage and mean ± standard deviation (SD), minimum and maximum values of items. By grouping as Estimated and actual DIF I, acceptable DIF I, and other DIF I (easy DIF I + difficult DIF I), one-to-one matching was evaluated with Mc neamer chi-square test. DIF I and DI values among the groups formed according to the distractor activity were compared using the t-test, and the effect size was calculated. A p-value of <0.05 was considered statistically significant.

Ethical Approval: The study was conducted after receiving approval from Non-interventional Research Ethics Committee of Dokuz Eylül University (Decision Date: 11.03.2023, No:2023/02-11).

Table 1. Distribution of items according to their DIF I, DI and NFD* (n=688)

		Number of items	%
DIF I	≤ 0.29 (too difficult)	69	10.0
	0.3 - 0.7 (acceptable)	414	60.2
	≥ 0.7 (too easy)	205	29.8
	0.5-0.6 (ideal)	327	47.5
DI	≤ 0.2 (poor)	183	26.6
	0.21-0.24 (acceptable)	57	8.3
	0.25-0.34 (good)	148	21.5
	≥0.35 (excellent)	300	43.6
NFD	0 NFD	207	30.1
	1 NFD	166	24.1
	2 NFD	154	22.4
	3NFD	104	15.1
	4 NFD	57	8.3

*DIF I: Difficulty Index, DI: Discrimination Index, NFD:Non-Functional Distractors

***Data usage permission**:* The data used in the research were obtained retrospectively from the item analysis results of the examinations conducted by the Dean's Office of Dokuz Eylül University Faculty of Medicine. The usage of the data has been granted permission by the Dean's Office of Dokuz Eylül University Faculty of Medicine (29.12.2022 / Document number: E-13511134-044[044]-470153). This permission document, along with other requested documents, was submitted to the Ethics Committee of Dokuz Eylül University for Non-interventional Studies. No personal data belonging to individuals were used, and no interventional procedures were performed.
This research dataset is accessible at https://doi.org/10.5281/zenodo.10461566.The anonymized data can be made accessible on request.

RESULTS

DIF I mean in post-evaluation item analysis was determined as 0.57±0.21 (min:0.06 max: 0.99). 60.2% (205 questions) of all MCQs were at the acceptable difficulty level, and 47.5% (327 questions) were at the ideal difficulty level (Table 1).
According to the difficulty estimates made by the MCQ authors while preparing the questions, 52.0% (358 questions) of the questions were labeled acceptable, 39.0% (268 questions) were labeled too easy, and 9.0% (62 MCQs) were labeled too difficult (Figure 1).

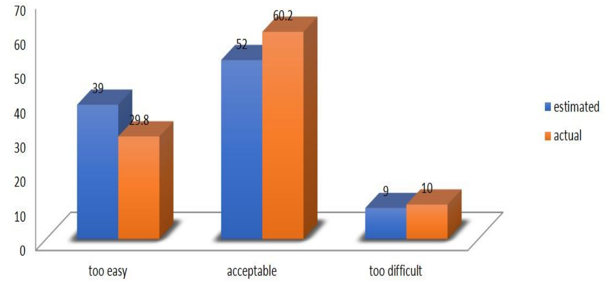

Figure 1. Distribution of items according to their estimated DIF I and actual DIF I (n=688) (%).

It was determined that there was a significant difference between the DIF I predicted by the MCQ authors and the actual DIF I groups ($x2_{mc\ neamer}$ =9.45, p=0.002) (Table 2).

The DI average for all items was 0.31 ± 0.17 (min: -0.19, max:0.84). In the grouping according to the DI level, the discrimination of 43.6% (300 questions) of the items was at a very good level, while the discrimination of 26.6% (183 questions) was very low (Table 1).

There were a total of 2752 distractors in 688 MCQs. 36.8% of the distractors (1014 distractors) were NFD. It was determined that all distractors worked in 30.1% of the MCQs (207 questions), and none of the distractors worked in 8.4% (57 MCQs) (Table 1).

In the comparison between DIF I and DI levels according to the operating status of the distractors, it was determined that there was a significant difference between the averages of the groups (p<0.001for all, η^2 =0.569 and 0.083, respectively) (Table 3).

DISCUSSION

In our study, it was determined that the DIF I value of 60.2% of all items was at an acceptable level, and 47.5% of them were at an ideal level. When we look at the estimated DIF I proportional distribution 52.0% acceptable, 39.0% too easy, 9.0% too difficult. There is a statistically significant difference between estimated and actual DIF ($x2_{mc\ neamer}$ =9.45, p=0.002). MCQs DI values, we found that 43.6% were at an excellent level and 21.5% were at a good level and it was found that all distractors worked in 30.1% of MCQs. When we compared the DIF I values of MCQs according to the number of NFDs, we saw that, as expected, as the number of NFDs increased, the DIF I value approached one, and the effect size was significant (p= 0.000, η^2= 0.569).

It is common to use tests consisting of MCQs for knowledge assessment in medical education. Item analyses are applied after the questions are used and provide valuable information about the test's overall quality and the MCQ's quality.

In the literature, many studies evaluated the results of item analysis of tests consisting of multiple-choice and single-correct answers applied in medicine and the health field. These studies generally aim to evaluate the items of only a single test (13, 14, 17,19). Unlike these studies, our research is based on the item analysis results of 688 of the 700 MCQs used in the Term 1 knowledge evaluation exams in the 2021-2022 academic year, which contained no information errors and were included in the question bank after the exams. These questions were used for the first time in the relevant exams.

Recognition/frequent use of MCQs by students has an impact on item analysis results. Therefore, we think it is important that all MCQs are used for the first time. However, we found no information on this subject in the studies we compared the results of.

In our study, it was determined that the DIF I value of 60.2% of all items was at an acceptable level, and 47.5% of them were at an ideal level. When our results are compared with similar studies in the literature, it is seen that our acceptable DIF I rate is lower than some studies (15, 18, 21, 22). However, approximately half (47.5%) of the 688 MCQs we evaluated in our research have an ideal DIF I value. This finding indicates that most questions within acceptable limits were stacked at an ideal level. The range defined as the acceptable limit is quite wide. MCQs in tests have a certain difficulty, and discrimination limit is a criterion that must be

Table 2. Distribution of MCQs according to Estimated and actual DIF I

		Actual DIF I						Total		
		Acceptable			Other					
		n	Row %	Column %	n	Row %	Column %	n	Row %	Column %
Estimated DIF I	**Acceptable**	226	63.1	54.6	132	36.9	48.2	358	100.0	52.0
	Other	188	57.0	45.4	142	43.0	51.8	330	100.0	48.0
Total		414	60.2	100.0	274	39.8	100.0	688	100.0	100.0

Table 3. Comparison of MCQs' DIF I and DIF Levels

		N	Mean	SD	Min.	Max.	F	p	Eta Squared(η^2)
DIF I	4 NFD	57	.8961	.04242	.81	.99	225.37	0.000	0.569
	3 NFD	104	.7550	.12403	.30	.90			
	2 NFD	154	.6353	.14244	.16	.84			
	1 NFD	166	.4868	.15984	.06	.78			
	0 NFD	207	.4127	.13096	.09	.71			
	Total	**688**	**.5722**	**.20576**	**.06**	**.99**			
DI	4 NFD	57	.1695	.08589	.00	.37	15.36	0.000	0.083
	3 NFD	104	.2905	.14291	-.06	.84			
	2 NFD	154	.3425	.15454	-.03	.77			
	1 NFD	166	.3449	.17484	-.10	.73			
	0 NFD	207	.3217	.16996	-.19	.66			
	Total	**688**	**.3146**	**.16502**	**-.19**	**.84**			

considered for the tests to serve their purpose and obtain valid and reliable measurement tools. Tests consisting of MCQs without appropriate values can also impact student exam success, both in terms of failure and passing (5, 8, 23, 24). Therefore, we think that the ideal level should be tried to be achieved rather than the acceptable level, and these questions should be given priority in the selection for the question bank.

MCQ writers need to consider these criteria while preparing their questions and try to prepare questions suitable for the ideal DIF I and DF level, free from spelling and editing errors. Our faculty has been providing training on MCQ preparation and item analysis for many years. However, as a common behavior, there are problems in complying with existing guidelines or training materials, as described in the literature (8, 24 - 29). Hence, MCQs in our Faculty are reviewed for item writing flaws by a measurement evaluator other than the question authors before they are used in the tests. Detected errors are corrected. In this way, the effect of common errors carried by MCQs is tried to be minimized. In the literature, in similar studies on item analysis, no information was found indicating that MCQs were reviewed/corrected by measurement and evaluator before being used. In the study of Ali and Ruit, based on the item analysis results, the results obtained in the subsequent use of the MCQs reviewed in terms of item writing flaws and NFD were evaluated (30). Research that provides the opportunity to demonstrate the effect of the regulations by comparing them with the item analysis results of equivalent exams consisting of MCQs without final adjustments will allow us to evaluate the real impact of the intense effort given.

When determining the difficulty of a question, the MCQ writer needs to consider the level of the learning goal related to the question and the cognitive level of the question, which is often overlooked. Our study evaluates the consistency of the item difficulty level predicted by the MCQ author with the actual difficulty level. When we look at the estimated DIF I proportional distribution (52.0% acceptable, 39.0% too easy, 9.0% too difficult), it can be thought that question writers generally tend to prepare MCQs at an acceptable difficulty level. However, when we examined the distribution of actual DIF I values, it was determined that there are proportional differences in the DIF I distribution compared to the estimated DIF I, within very easy and acceptable limits. In a one-to-one comparison, it was determined that only half of the MCQs, which were within acceptable limits according to the actual DIF I, were labeled in the same way by the question authors. There is a statistically significant difference between estimated and actual DIF ($x2_{mc\ neamer}$ =9.45, p=0.002). This situation is most likely because MCQ authors label DIF I without giving it much thought/care. It may also be effective that students do not have a realistic prediction about their knowledge level or that the MCQs are prepared by an assessment evaluator other than the author before the exams.

In our study, when we grouped the MCQs according to their DI values, we found that 43.6% were at an excellent level and 21.5% were at a good level. In our study, the excellent DI level is lower than that of Uddin et al.'s study; however, the number of evaluated MCQs specifically given in percentage was not clearly defined in the article (20). Our excellent DI level question rate was at a higher level than similar studies, except for the study by Rao et al. The high rate of questions with ideal DIF I and excellent DI levels can also be considered an important clue about exam reliability (15, 18, 21, 22).

The presence of NFD is one of the factors affecting the quality of MCQ. In our research, it was found that all distractors worked in 30.1% of MCQs. In 8.4% of MCQs, all distractors were non-functional. These MCQs may have irrelevant distractors, the cognitive level may be very low, or they may indicate rare situations in which all students achieve correct learning. However, the rate of these MCQs is quite low. The results of the study are consistent with the study of Kumar et al., where the rate of MCQs with all distractors being functional was found to be 33%, and the rate of MCQs with all distractors being non-functional was found to be 2% (22). However, Bhattacherjee et al., found in their study that the rate of MCQ with all distractors working was 13.33%, and the rate of MCQ with all distractors being non-functional was 16.67% (21).

We use five-choice MCQs with one correct answer in our exams. However, in studies conducted on medical and health education exams related to item analysis, it is observed that MCQs are arranged with one correct answer and four options. Studies are showing that the number of options in the MCQ being less than five does not have a significant effect on the DIF I and DI values or that increasing the number of distractors in the MCQ does not have a positive effect (31-37). Kheyami et al., state that using an MCQ with four options may be better than using an MCQ with five options (38). Rodríguez, reported that removing the least functional distractor did not have a negative effect on DIF I, while the remaining ones may have a positive effect on DI with higher selection frequency (39).

When we compared the DIF I and DI values of MCQs according to the number of NFDs, we saw that, as expected, as the number of NFDs increased, the DIF I value approached one, and the effect size was significant ($p= 0.000$, $\eta^2= 0.569$). In our analysis of the DI values of MCQs according to the number of NFDs, we revealed that although we found a significant difference, the effect size was very low ($p= 0.000$, $\eta^2=0.083$). This finding is consistent with the results of the studies of Rodriguez, Hingorjo and Jaleel (39, 40). The NFD number affects both the DIF I and DI value of the problem. This effect is worth considering, especially on DIF I. It is a difficult task to prepare MCQs with DIF I and DI values that can be considered ideal for assessments in medical education, and as the number of options increases, more editing effort and time is required for the authors. Based on this finding, which is compatible with the literature, we think that it would be appropriate to prepare MCQs with four well-constructed options.

CONCLUSION

In our study, we found that MCQs had mostly ideal DIF I values, but the MCQ authors were not very accurate in their DIF I estimations. In our study, we found that the DI value of approximately half of the questions was excellent. We think that questions with ideal DIF I and DI values will provide a realistic evaluation opportunity. While the number of non-functional distractors has a significant effect on the DIF of a question, we found that it has a significant but small effect on the DI. The effect of the number of NFDs on DIF I and DI suggested that questions with four options could be used instead of having difficulty and making mistakes while trying to create questions with five options. We did not have the opportunity to evaluate the effect of reviewing MCQs in terms of item writing flaws before being used in exams on item analysis values, but we think that studies on this subject will be valuable. It is important for MCQ authors to use the item analysis results in the question bank records as a guide for their development and to avoid repeating the same mistakes in the questions they have just prepared to create reliable tests with high measurement values.

Assessing the item analyses of numerous MCQs, particularly examining the impact of NFDs on DIF I and DI, and comparing question writers' difficulty predictions with the actual difficulty encountered constitute the strong aspects of this study. On the other hand, the failure to evaluate the impact of having questions reviewed by someone other than the question writers before usage constitutes a weakness of this study.

Acknowledgements: None.

Author Contribution: The author SK worked independently in conceptualizing and designing the study, collecting and analyzing data, drafting, and finalizing the manuscript.

Conflict of interest: The author declares that there is no potential conflict of interest to be disclosed.

Ethical Approval: The study was approved by Non-interventional Research Ethics Committee of Dokuz Eylül University (Decision Date: 11.03.2023, No:2023/02-11).

Funding: No funding sources.
Peer-review: Externally peer-reviewed.

REFERENCES

1. Carneson J, Delpierre G, Masters K. Designing, and managing multiple choice questions. 2nd ed. 2016; pp. 3–6. University of Cape Town. Available from: https://www.researchgate.net/publication/309263856_Designing_and_Managing_Multiple_Choice_Questions_2nd_Ed=channel=doi&linkId=58074fef08ae03256b783474&showFulltext=true .
2. 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;4:23.
3. Gierl MJ, Lai H, Turner SR. Using automatic item generation to create multiple-choice test items. Med Educ 2012;46(8):757-765.
4. Palmer EJ, Devitt PG. Assessment of higher order cognitive skills in undergraduate education: modified essay or multiple choice questions? Research paper. BMC Med Educ 2007;7:49.
5. 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(2):198-206.
6. Tavakol M & Dennick R. Post-examination analysis of objective tests. Medical Teacher 2011;33(6):447–458.
7. Case SM, Swanson DB. Constructing written test questions for the basic and clinical sciences. 3rd ed. Philadelphia: National Board of Medical Examiners; 2000. Availablefrom:https://www.researchgate.net/publication/242759434_Constructing_Written_Test_Questions_For_the_Basic_and_Clinical_Sciences
8. Downing SM. The effects of violating standard item writing principles on tests and students: the consequences of using flawed test items on achievement examinations in medical education. Adv Health Sci Educ Theory Pract. 2005;10(2):133-143.
9. Haladyna TM, Downing SM & Rodriguez MC A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment, Applied Measurement in Education 2002;15(3):309-333.
10. Medical Council of Canada (MCC). Guidelines for the Development of Multiple-Choice Questions. Ottawa, ON: MCC 2010. Available from: https://mcc.ca/media/Multiple-choice-question-guidelines.pdI
11. Paniagua MA, Swygert K A editors. Constructing written test questions for the basic and clinical sciences (2016). Available from: https://www.bumc.bu.edu/busm/files/2018/10/NBME-Constructing-Written-Test-Questions.pdI
12. Sutherland K, Schwartz J, Dickison P. Best Practices for Writing Test Items. Journal of Nursing Regulation 2012;3(2):35-39.
13. 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(6):1876-81.
14. Date AP, Borkar AS, Badwaik RT, Siddiqui RA, Shende TR, & Dashputra AV. Item analysis as tool to validate multiple choice question bank in pharmacology. International Journal of Basic & Clinical Pharmacology 2019;8(9):1999–2003.
15. Rehman A, Aslam A & Hassan SH. Item analysis of multiple choice questions. Pakistan Oral & Dental Journal 2018;38(2): 291-293.
16. Gierl MJ, Bulut O, Guo Q & Zhang X. Developing, Analyzing, and Using Distractors for Multiple-Choice Tests in Education: A Comprehensive Review. Review of Educational Research 2017;87(6):1082–1116.
17. Hassan S & Hod R. Use of item analysis to improve the quality of single best answer multiple choice question in summative assessment of undergraduate medical students in Malaysia. Education in Medicine Journal 2017;9(3):33-5-43.
18. Rao C, Kishan Prasad H L, Sajitha K, Permi H, Shetty J. Item analysis of multiple choice questions: Assessing an assessment tool in medical students. Int J Educ Psychol Res 2016;2:201-4.
19. Kolte V. Item analysis of Multiple Choice Questions in Physiology examination. Indian Journal of Basic and Applied Medical Research; 2015;4(4):320-326.
20. Uddin I, Uddin İ, Rehman IU, Siyar M, Mehbob U. Item Analysis of Multiple Choice Questions in Pharmacology. J Saidu Med Coll Swat 2020;10(2):128-13.
21. Bhattacherjee S, Mukherjee A, Bhandari K, Rout AJ. Evaluation of Multiple-Choice Questions by Item Analysis, from an Online Internal Assessment of 6th Semester Medical Students in

a Rural Medical College, West Bengal. Indian J Community Med. 2022;47(1):92-95.

22. 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(1):85–89.
23. Masters JC, Hulsmeyer BS, Pike ME, Leichty K, Miller MT & Verst AL. Assessment of multiple-choice questions in selected test banks accompanying text books used in nursing education. The Journal of Nursing Education 2001;40(1):25–32.
24. Walsh K. Advice on writing multiple choice questions (MCQs). BMJ 2005;330: 25 - 27.
25. Costello E, Holland JC, Kirwan C. Evaluation of MCQs from MOOCs for common item writing flaws. BMC Res Notes. 2018;11(1):849.
26. Downing SM. Construct-irrelevant variance and flawed test questions: Do multiple-choice item-writing principles make any difference? Academic medicine: journal of the Association of American Medical Colleges 2002;77(10):103–104.
27. Gupta P, Meena P, Khan AM, Malhotra RK & Singh T. Effect of Faculty Training on Quality of Multiple-Choice Questions. International journal of applied & basic medical research 2020; 10(3):210–214.
28. Huang Yi-Min, Trevisan M, Storfer A. The Impact of the “all-of-the-above” Option and Student Ability on Multiple Choice Tests. International Journal for the Scholarship of Teaching and Learning 2007;1(2):11.
29. Scott KR, King AM, Estes MK, Conlon LW, Jones JS & Phillips AW. Evaluation of an Intervention to Improve Quality of Single-best Answer Multiple-choice Questions. The Western Journal of Emergency Medicine 2019;20(1):11–14.
30. Ali SH & Ruit KG. The Impact of item flaws, testing at low cognitive level, and low distractor functioning on multiple-choice question quality. Perspectives on medical education 2015; 4(5): 244–251.
31. Al-Lawama M & Kumwenda B. Decreasing the options' number in multiple choice questions in the assessment of senior medical students and its effect on exam psychometrics and distractors' function. BMC Medical Education 2023;23(1):212.
32. 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):635.
33. 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):252.
34. Pawade YR & Diwase DS. Can Item Analysis of MCQs Accomplish the Need of a Proper Assessment Strategy for Curriculum Improvement in Medical Education? i-manager’s Journal of Educational Technology 2016;13(1):44-53.
35. Rogausch A, Hofer R & Krebs R. Rarely selected distractors in high stakes medical multiple-choice examinations and their recognition by item authors: a simulation and survey. BMC medical education 2010;10:85.
36. 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: 40.
37. Rahma A, Shamad M, Idris ME, Elfaki O, Elfakey W, Salih KM. Comparison in the quality of distractors in three and four options type of multiple-choice questions. Adv Med Educ Pract. 2017;8:287–91.
38. 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(1):68–74.
39. Rodriguez MC. Three Options Are Optimal for Multiple-Choice Items: A Meta-Analysis of 80 Years of Research. Educational Measurement: Issues and Practice 2005; 24(2):3-13.
40. Hingorjo MR & Jaleel F. Analysis of one-best MCQs: the difficulty index, discrimination index and distractor efficiency. JPMA. The Journal of the Pakistan Medical Association 2012;62(2): 142–147.