SANAMED
2024; 19(2): 163–171
ISSN-1452-662X

DOI: 10.5937/sanamed0-50691
UDK: 37.091.27:61-057.875
ID: 149843721
Original article

ITEM ANALYSIS OF MULTIPLE-CHOICE QUESTIONS IN AN UNDERGRADUATE SURGERY COURSE-AN ASSESSMENT OF AN ASSESSMENT TOOL

Al Sulaim Sulaiman Lamees, Salati Ahmad Sajad

Qassim University, College of Medicine, Department of Surgery, Unaizah, Kingdom of Saudi Arabia

Primljen/Received 27. 04. 2024. Prihvaćen/Accepted 06. 06. 2024. Published online first: 24. 06. 2024.

Abstract: **Introduction:** In the field of medical education, multiple-choice questions (MCQs) represent the most commonly utilized method of assessment. It is necessary to analyze the assessment results through item analysis to ensure the quality is appropriate. This study evaluated the quality of the MCQs utilized for summative evaluation of the students in the General Surgery Course conducted in the year 2023-24, at the College of Medicine (Unaizah), Qassim University, Saudi Arabia.

Methods: Using a number of established parameters for item analysis, the study evaluated the multiple-choice questions for difficulty, discrimination power, and quality of distractors.

Results: The quality of the questions varied. The means of the facility index, discrimination index, discriminative efficiency, and distractor efficiency were, in order, 76.31%, 0.28, -0.7743, and 32%.

Conclusion: Item analysis is a crucial technique for evaluating the quality of MCQs. There were multiple defects in the MCQs used in summative assessments, revealing the scope for further improvement in future courses. It is important to plan faculty development events often to impart knowledge and skills related to creating MCQs that are valid, reliable, and of high quality.

Keywords: Assessment**,** Item analysis, Multiple-choice questions, Facility index, Discriminative index, Distractor efficiency, Non-functional distractor, Distractor analysis.

INTRODUCTION

Assessment is an essential process used in medical education to evaluate a student's clinical competency and capacity to meet predetermined learning objectives (1). The assessment tools are devised to enhance students' comprehension and learning while enabling them to identify their areas of weakness (2). From the teachers' perspective, the assessment outcomes provide evidence to uphold or alter the educational objectives and pedagogical approaches (3). To satisfy various evaluation demands, a wide variety of assessment methods are available, including multiple-choice questions (MCQs), the Objective Structured Clinical Exam (OSCE), short answer questions (SAQs), modified essay questions (MEQs), extended matching questions (EMQs), mini-clinical evaluation exercises (Mini-CEX), direct observation of procedural skills (DOPS), and viva-voce.

Whether an assessment is being used for formative (diagnosis, feedback, and improvement) or summative (promotion and certification) reasons will determine which method is best. There are several recognized attributes of assessment tools, including affordability, practicality, validity, reliability, and educational impact. Regardless of the reason, a single assessment method cannot evaluate every competency domain, and hence, a range of assessment methods is needed, allowing the benefits of one to offset the drawbacks of another (4).

MCQs are the most widely used assessment tool in medical education worldwide, and they come in many types. Type-A MCQ comprises two parts: a stem that states the question or lead-in; and a set of alternatives, or potential answers, which include a key, which is the best response to the question, and several distractors (three to four), which are reasonable but incorrect responses. MCQs are supposed to assess all levels of teaching objectives within Bloom's taxonomy of learning, including knowledge recall, comprehension, application, analysis, synthesis, and evalua-

tion. Making an excellent multiple-choice test (MCQ) can be difficult and time-consuming. Nevertheless, because answers to MCQs can be evaluated quickly and precisely, this approach is usually chosen over other evaluation tools due to its objectivity and reduction of human bias (5,6,7). A good-quality MCQ has key answers distributed throughout the test, thereby minimizing placement bias, and the key and distractors must be similar in length, style, and grammatical form. To audit the quality of the assessment, the MCQs are analyzed by a process termed item analysis (8).

The aim of this study is to evaluate the quality of the multiple-choice questions (MCQs) utilized in the surgery block during the academic year 2023-24 for the Year-4 undergraduate medical students at the College of Medicine, Qassim University, Saudi Arabia, as part of their summative assessment.

MATERIAL AND METHODS

A cross-sectional study was conducted at the Department of Surgery, College of Medicine (Unaizah branch), Qassim University, Saudi Arabia, during the academic session 2023-24. The college was established in 2012 and has adopted a team-based learning (TBL) curriculum for undergraduate medical training, leading to the award of an MD degree upon successful completion. The assessment is based on the curricular contents related to the intended learning objectives, and multiple-choice questions (MCQs) are one of the tools utilized for assessment in the surgery course.

The study analyzed the results of the MCQs used in the summative assessment of 38 male students in Year 4 of the medical undergraduate course, conducted at the end of 8 weeks of teaching. The MCQs were analyzed for:

i. Their level of difficulty, measured by the facility index (P),

ii. Power of discrimination, measured by the discrimination index (DI), and

iii. Distractor analysis, measured by distractor efficiency (DE).

There were 50 items, and the time allotted for each item was 2 minutes (total time: 100 minutes). Each item was of the one-best type, having a single stem and five answer options, one of them being correct and the other four being 'distractors.' The test was conducted paperless on a computer, and students were required to log in through their university-allotted usernames and passwords. Every correct response was awarded 1 mark, and there was neither a mark nor a negative mark for any blank or incorrect response. Thus, the maximum and minimum attainable scores for the test were 50 and 0, respectively. The test was criterion-referenced, and passing standards were expressed in absolute terms, with a passing score of 60%. For the test, 70% of MCQs were newly constructed, and 30% were taken after modification from the question bank created from the tests conducted over the previous five years.

The scores of the students were arranged in descending order, and then the DI was determined using Kelley's technique, which takes into account the difference between the scores achieved by 27% of students on the higher side (high achievers) and those of the 27% on the lower side (low achievers). Calculations of the values of P and DI were undertaken by the application of the formulae depicted in Table 1.

Table 1. *Formulae for calculation of facility index (P) & discrimination index (DI) of MCQs*

Facility Index & Discrimination Index Calculations	
Facility Index (P) = [(H+L) / N] x 100	**Inference** *: Less than 30%, the MCQ is very difficult.* *Greater than 70%, the MCQ is easy.* *30% to 70%, the MCQ lies within an acceptable range.*
Discrimination Index (DI) = 2 x [(H-L) / N]	**Inference** : The DI measures the differences obtained in correct responses between the higher achievers and the lower achievers. The calculated value ranges between 0 and 1. The higher the DI, the more the test item can discriminate better between students with higher and lower test scores. Accordingly, if the value is: ***1.*** *0.19 or less, the MCQ has poor discrimination.* ***2.*** *Between 0.2 and 0.29, the MCQ has acceptable discrimination.* ***3.*** *Between 0.3 and 0.4, the MCQ has good discrimination.* ***4.*** *Greater than or equal to 0.4, the MCQ has excellent discrimination.*
• *N is the sum total of the students in both high and low groups.* • *H and L respectively stand for the number of correct responses in the high and low groups.* *MCQs that attained a P between 30 - 70 and DI > 0.24 were termed* ***'ideal'****.*	

The distractors in the MCQ are measures of its functioning. When a distractor is chosen by more than 5% of the examinees, it is considered a functioning distractor (FD), and if chosen by less than 5%, it is termed a non-functioning distractor (NFD). On the basis of the number of NFDs in an MCQ, distractor efficiency (DE) ranged from 0 to 100%, as shown in Table 2.

Table 2. *Relationship of Distractor Efficiency (DE) with the Number of Non-Functional Distractors (NFDs)*

Non-Functional Distractors (NFDs)	Distractor Efficiency (DE)
4	0
3	25
2	50
1	75
0	100

RESULTS

The item analysis of the surgery course multiple-choice question (MCQ) exam revealed important findings regarding the quality of the assessment tool. There were 50 five-option MCQs, with a single key and four distractors in each. Thirty-two students appeared in the examination and attained grades ranging from 16.3 to 50 out of the maximum of 50, with a mean score of 40.47 ± 8.08, as shown in Table 3. There were only 12 MCQs (24%) with the facility index (FI) lying in the desired range of 30–70%. Thirty-five MCQs had a facility index above 70%, hence qualifying as easy to solve. The examination demonstrated an adequate level of discrimination overall between high- and low-performing students, as 40 (80%) MCQs had a positive discrimination index (DI). However, since only 21 (42%) MCQs had good or excellent discrimination, there is significant scope for improvement.

Tables 4 and 5 show the item analysis of distractors for MCQs 1-25 and MCQs 26-50 respectively, while Table 6 displays their overall psychometric analysis. This analysis pointed towards a deficiency in the design of the distractors, as depicted in the following data:

i. Seventeen (34%) MCQs had 4 non-functioning distractors (NFDs) and hence, a distractor efficiency (DE) percentage of 0%.

ii. Twelve (24%) MCQs had 3 NFDs and hence, a DE percentage of 25%.

iii. Twelve (24%) MCQs had 2 NFDs and hence, a DE percentage of 50%.

iv. Eight (16%) MCQs had 1 NFD and hence, a DE percentage of 75%.

v. One (2%) MCQ had no NFD and hence, a DE percentage of 100 %.

Table 3. *Item analysis report*

Serial Number	Parameter	Results
1.	Exam Median score	42
2.	Exam Average Score (Mean)	40.47
3.	Standard Error of Measurement	1.4
4.	Standard Deviation (SD)	8.08
5.	Minimum Score	16.3
6.	Maximum Score	50
7.	Total Difficulty	76.31%
8.	MCQs with Facility Index < 30%	3
9.	MCQs with Facility Index > 70%	35
10.	MCQs with –ve discrimination	5
11.	MCQs with zero discrimination	5
12.	MCQs with +ve discrimination Index < 0.19	13
13.	MCQs with +ve discrimination Index = 0.2-0.29	6
14.	MCQs with +ve discrimination Index = 0.3-0.39	3
15.	MCQs with +ve discrimination Index >=0.4	18

Table 4. *Distractor Analysis (MCQs 1-25)*

Serial	Key Answer	Number and percentage of students selecting the options					NFDs
		A	B	C	D	E	
1.	*C*	1 (3.13%)	21 (65.63%)	5 (15.63%)	3 (9.38%)	2 (6.25%)	1
2.	*D*	0 (0.00%)	4 (12.50%)	0 (0.00%)	23 (71.88%)	5 (15.63%)	2
3.	*B*	1 (3.13%)	16 (50.00%)	10 (31.25%)	2 (6.25%)	3 (9.38%)	1
4.	*B*	0 (0.00%)	32 (100.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	4
5.	*B*	0 (0.00%)	23 (71.88%)	5 (15.63%)	2 (6.25%)	2 (6.25%)	1
6.	*C*	1 (3.13%)	0 (0.00%)	30 (93.75%)	1 (3.13%)	0 (0.00%)	4
7.	*C*	7 (21.88%)	15 (46.88%)	10 (31.25%)	0 (0.00%)	0 (0.00%)	2
8.	*A*	31 (96.88%)	1 (3.13%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	4
9.	*A*	29 (90.63%)	1 (3.13%)	1 (3.13%)	1 (3.13%)	0 (0.00%)	4
10.	*E*	2 (6.25%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	30 (93.75%)	3
11.	*C*	3 (9.38%)	2 (6.25%)	24 (75.00%)	3 (9.38%)	0 (0.00%)	1
12.	*B*	0 (0.00%)	32 (100.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	4
13.	*B*	0 (0.00%)	30 (93.75%)	0 (0.00%)	1 (3.13%)	1 (3.13%)	4
14.	*D*	2 (6.25%)	0 (0.00%)	1 (3.13%)	29 (90.63%)	0 (0.00%)	3
15.	*B*	0 (0.00%)	30 (93.75%)	0 (0.00%)	0 (0.00%)	2 (6.25%)	3
16.	*A*	0 (0.00%)	0 (0.00%)	1 (3.13%)	0 (0.00%)	31 (96.88%)	4
17.	*B*	0 (0.00%)	29 (90.63%)	3 (9.38%)	0 (0.00%)	0 (0.00%)	3
18.	*E*	10 (31.25%)	0 (0.00%)	2 (6.25%)	5 (15.63%)	15 (46.88%)	1
19.	*D*	0 (0.00%)	0 (0.00%)	0 (0.00%)	31 (96.88%)	1 (3.13%)	4
20.	*E*	2 (6.25%)	3 (9.38%)	0 (0.00%)	0 (0.00%)	27 (84.38%)	2
21.	*A*	25 (78.13%)	2 (6.25%)	1 (3.13%)	4 (12.50%)	0 (0.00%)	2
22.	*A*	31 (96.88%)	0 (0.00%)	1 (3.13%)	0 (0.00%)	0 (0.00%)	4
23.	*D*	5 (15.63%)	0 (0.00%)	0 (0.00%)	27 (84.38%)	0 (0.00%)	3
24.	*D*	0 (0.00%)	1 (3.13%)	1 (3.13%)	22 (68.75%)	8 (25.00%)	3
25.	*C*	8 (25.00%)	0 (0.00%)	22 (68.75%)	0 (0.00%)	2 (6.25%)	2

Table 5. *Distractors analysis (MCQs 26-50)*

Serial	Key Answer	Number and percentage of students selecting the options					NFDs
		A	B	C	D	E	
26.	*E*	0 (0.00%)	0 (0.00%)	0 (0.00%)	0 (0.00%)	32 (100.00%)	4
27.	*B*	2 (6.25%)	1 (3.13%)	20 (62.50%)	8 (25.00%)	1 (3.13%)	2
28.	*B*	30 (93.75%)	0 (0.00%)	0 (0.00%)	2 (6.25%)	0 (0.00%)	3
29.	*D*	0 (0.00%)	13 (40.63%)	0 (0.00%)	17 (53.13%)	2 (6.25%)	2
30.	*B*	1 (3.13%)	30 (93.75%)	0 (0.00%)	0 (0.00%)	1 (3.13%)	4
31.	*B*	5 (15.63%)	25 (78.13%)	2 (6.25%)	0 (0.00%)	0 (0.00%)	2
32.	*B*	2 (6.25%)	22 (68.75%)	2 (6.25%)	5 (15.63%)	1 (3.13%)	1
33.	*D*	11 (34.38%)	1 (3.13%)	1 (3.13%)	19 (59.38%)	0 (0.00%)	3
34.	*B*	0 (0.00%)	31 (96.88%)	1 (3.13%)	0 (0.00%)	0 (0.00%)	4
35.	*E*	0 (0.00%)	2 (6.25%)	1 (3.13%)	1 (3.13%)	28 (87.50%)	3
36.	*B*	14 (43.75%)	10 (31.25%)	3 (9.38%)	3 (9.38%)	2 (6.25%)	0
37.	*A*	29 (90.63%)	0 (0.00%)	1 (3.13%)	2 (6.25%)	0 (0.00%)	3

38.	*D*	22 (68.75%)	2 (6.25%)	0 (0.00%)	8 (25.00%)	0 (0.00%)	2
39.	*C*	1 (3.13%)	0 (0.00%)	31 (96.88%)	0 (0.00%)	0 (0.00%)	4
40.	*A*	13 (40.63%)	0 (0.00%)	0 (0.00%)	17 (53.13%)	2 (6.25%)	2
41.	*C*	0 (0.00%)	1 (3.13%)	31 (96.88%)	0 (0.00%)	0 (0.00%)	4
42.	*C*	0 (0.00%)	0 (0.00%)	28 (87.50%)	2 (6.25%)	2 (6.25%)	2
43.	*D*	2 (6.25%)	3 (9.38%)	4 (12.50%)	23 (71.88%)	0 (0.00%)	1
44.	*C*	0 (0.00%)	1 (3.13%)	23 (71.88%)	5 (15.63%)	3 (9.38%)	2
45.	*B*	0 (0.00%)	30 (93.75%)	0 (0.00%)	1 (3.13%)	1 (3.13%)	4
46.	*C*	1 (3.13%)	5 (15.63%)	26 (81.25%)	0 (0.00%)	0 (0.00%)	3
47.	*C*	2 (6.25%)	2 (6.25%)	24 (75.00%)	3 (9.38%)	1 (3.13%)	1
48.	*D*	1 (3.13%)	3 (9.38%)	0 (0.00%)	28 (87.50%)	0 (0.00%)	3
49.	*D*	0 (0.00%)	1 (3.13%)	0 (0.00%)	30 (93.75%)	1 (3.13%)	4
50.	*D*	0 (0.00%)	0 (0.00%)	0 (0.00%)	32 (100.00%)	0 (0.00%)	4

Table 6. *Psychometric Analysis of Multiple-Choice Questions*

Serial Number	***Facility Index***	***Discrimination Index***	***Discriminative Efficiency***	***Distractor Efficiency***	***Serial Number***	***Facility Index***	***Discrimination Index***	***Discriminative Efficiency***	***Distractor Efficiency***
1	15.63%	0.17	33.33%	75%	***26***	100.00%	0%	0%	0%
2	71.88%	0.39	48.39%	50%	***27***	62.50%	0.19	22.94%	50%
3	*50.00%*	*0.17*	*2.16%*	*75%*	***28***	*0.00%*	*0%*	*0%*	*25%*
4	*100.00%*	*0%*	*0%*	*0%*	***29***	*53.13%*	*0.54*	*68.83%*	*50%*
5	*71.88%*	0.67	83.85%	75%	***30***	93.75%	0.22	37.10%	0%
6	93.75%	0.78	13.51%	0%	***31***	78.13%	0.48	61.15%	50%
7	31.25%	-0.04	-5.68%	50%	***32***	68.75%	0.23	28.02%	75%
8	96.88%	-0.09	-17.37%	0%	***33***	59.38%	0.56	69.26%	25%
9	90.63%	0.52	79.26%	0%	***34***	96.88%	0.04	7.34%	50%
10	93.75%	-0.01	-2.21%	25%	***35***	87.50%	0.34	48.94%	25%
11	75.00%	0.28	34.14%	75%	***36***	31.25%	0.15	21.09%	50%
12	100.00%	0%	0%	0%	***37***	90.63%	0.19	28.06%	25%
13	93.75%	0.55	95.91%	0%	***38***	68.75%	0.20	25.40%	50%
14	90.63%	0.52	79.26%	25%	***39***	96.88%	-0.06	-11.20%	0%
15	93.75%	0.55	95.91%	25%	***40***	53.13%	0.59	75.89%	50%
16	96.88%	-0.06	-11.20%	0%	***41***	96.88%	0.04	7.34%	0%
17	90.63%	0.52	81.68%	25%	***42***	87.50%	0.57	84.66%	50%
18	46.88%	0.15	20.22%	75%	***43***	71.88%	0.11	13.76%	75%
19	96.88%	0.04	7.34%	0%	***44***	71.88%	0.45	56.41%	50%
20	84.38%	0.45	61.34%	50%	***45***	93.75%	0.48	83.63%	0%
21	78.13%	0.47	59.38%	50%	***46***	81.25%	0.30	38.96%	25%
22	96.88%	0.49	100.00%	0%	***47***	9.38%	0.19	44.99%	75%
23	84.38%	0.26	34.82%	25%	***48***	87.50%	0.63	92.33%	25%
24	68.75%	0.13	15.79%	25%	***49***	93.75%	0.29	50.90%	0%
25	68.75%	0.19	23.18%	50%	***50***	100.00%	0%	0%	0%

Mean Values: Facility Index 76.31%; Discrimination Index 0.28; Discriminative Efficiency -0.774; Distractor Efficiency 32%.

DISCUSSION

Assessment is an important component of medical education that, when executed appropriately, indicates whether or not students have achieved the intended learning goals (9). Instructional and curriculum modifications are also influenced by the assessment results (10). A variety of methods are available to assess knowledge, but multiple-choice questions (MCQs) are currently one of the most popular options because of their ease of use, impartiality, consistency, simplicity in administration, and capacity to cover a larger range of subject matter. MCQs can reveal information on students' comprehension, knowledge, and analytical abilities, allowing for the identification of both the strengths and flaws in their grasp of a subject. It has been shown that, when properly constructed, MCQs can assess higher-order cognitive domains such as synthesis and application, in addition to discriminating between students' individual abilities (11). On the other hand, if a test has a greater rate of errors, it tends to be less reliable and penalizes participants (12). As a result, it's critical to have robust distractors, rationally sound keys, and an effective stem that integrates the various learning levels and the directive verbs that go along with them within each learning domain (8).

Even though creating multiple-choice questions (MCQs) seems easy, it takes a lot of work and time to design them correctly, especially for faculty members who have never undergone dedicated training in assessment methodology (13). MCQs present a number of design challenges, such as confusing stems, multiple correct answers, contentious answers, give-away keys, poor distractors, and a high likelihood of predicting the right responses. Nonetheless, item analysis can offer the relevant data required to raise the caliber and efficacy of MCQs. As shown in Tables 3-5, our study identified a number of weaknesses regarding the degree of difficulty, the quality of distractions, and discrimination. The number of non-functional distractors (NFDs) in MCQs was of particular concern. The negative and zero discrimination index implied, contrary to expectations, that low-performing students accurately recognized the MCQ's key more often than or as frequently as good performers. This could be the result of an incorrect key, imprecise item phrasing, or possibly deficient student preparation (13).

To maintain fairness and the integrity of the test, various actions were initiated as per the academic regulations to neutralize the impact of defective items with a negative or zero discrimination index, including dropping out or modification of the questions. The shortcomings identified by item analysis have been discussed in many other published studies, and a high percentage of items with writing flaws (IWFs) and non-functional distractors (NFD) have been demonstrated (7, 14). Chauhan et al. (15) reported 1, 2, 3, and 4 NFDs with rates of 7.69%, 30.77%, 60.00%, and 1.54%. The percentage of 0, 1, 2, and 3 NFDs was 65.00%, 25.00%, 10.00%, and 0.00%, respectively, in Patel's (16) item analysis, where the MCQs included only three distractors. Similarly, Mahjabeen et al. (17) reported the figures at 25.00%, 46.00%, 25.00%, and 5.00%, respectively. The distractor effectiveness of the MCQs in this study is 32%, which is significantly lower than the values in the literature. Rao et al. (18), Gajjar et al. (19), and Patel (16) have achieved values as high as 90%, 89.6%, and 84.9%, respectively. In a study by Lama et al. (20), 43.3% of MCQs used in summative assessment of undergraduate dental students had a poor discrimination index (≤ 0.2), and distractor efficiency was 100 in only 6%. About 37% of the MCQs were either very difficult or very easy and hence inappropriate. The item analysis of MCQs used in the assessment of ophthalmology block in undergraduate medical courses found 50% to be defective, with bad stems or distractors being the prime culprits (21).

In a recent study by Baste (22), the correlation of the actual difficulty level of MCQs in the physiology block of undergraduate medical students as derived by item analysis was compared with the difficulty level as perceived by the faculty. It was found that the correlation between the actual and perceived difficulty level was poor, even though the enrolled faculty in the study were experienced. It was concluded that mere experience does not assure the accuracy of the perceived difficulty level and that item analysis needs to be properly conducted after every MCQ-based assessment. The study further revealed underutilization of item analysis of MCQs, and the reasons provided by the faculty included lack of motivation, involvement in difficult calculations, staff shortage, and lack of skills.

In a multidisciplinary and integrated curriculum, designing MCQs consumes a lot of time and is usually a difficult undertaking. Hence, there are recommendations about the disposal of MCQs found defective in item analysis. Bhat and Prasad (21) suggest such MCQs be analyzed to detect the item writing flaws and then optimized into a viable question. In their study, 16 out of 40 (40%) MCQs were either very easy or very difficult on the basis of item analysis, and after proper edits, 15 MCQs were salvaged for entry into the question bank.

The study's findings concur with the recommendation in the literature (21, 22) that faculty members should more frequently attend faculty development programs aimed at improving their ability to construct

excellent multiple-choice questions and create viable question banks. Numerous studies have demonstrated that carefully designed, longitudinal faculty development workshops improve the writing abilities of multiple-choice questions (MCQs) in terms of discriminating and difficulty indices, as demonstrated by the cognitive levels of Bloom's taxonomy, decreased item writing errors, and increased functioning distractors (13, 23, 24, 25).

CONCLUSION

Properly constructed multiple-choice questions (MCQs) are an objective and reliable tool to assess the learning performance of students. Item analysis is an important activity that must be properly conducted after MCQ tests in order to assess the level of difficulty and their capacity to distinguish between good and weak students. The results of this analysis have the potential to identify the sections of the course material that require revision or adjustment. Item analysis identified an array of weaknesses with our multiple-choice questions that required rectification as per the rules. Regular faculty development activities are required so that the MCQ constructors know how to correctly interpret the item-analysis data and, on that basis, undertake meaningful steps to improve the quality of questions, thereby achieving the objective of holding valid, effective, and fair tests. Only the question items with a good difficulty index, acceptable discrimination power, and zero non-functional distractors should be utilized for student promotion and retained in the question bank for possible reuse.

Abbreviations

MCQ - Multiple-choice question
NFD - Non-functioning distractor
DE - Distractor efficiency
DI - Discrimination index
P - Facility Index
IWF -Item writing flaw

Acknowledgment: The authors express gratitude to the Examination Committee of the college for providing the relevant data.

Conflict of Interests: The authors declare that there is no conflict of interest related to this article.

Funding: There is no source of funds/sponsors for this study.

Authors' Contribution: Both authors have contributed equally in conceiving the project, analyzing the data, reviewing the literature, and drafting the manuscript.

Note: Artificial intelligence was not used in any way in the design of this study endeavor.

Sažetak

ANALIZA STAVKI PITANJA SA VIŠESTRUKIM IZBOROM NA DODIPLOMSKOM ISPITU IZ HIRURGIJE - PROCENA ALATA ZA OCENJIVANJE

Al Sulaim Sulaiman Lamees, Salati Ahmad Sajad

Univerzitet Qassim, Medicinski fakultet, Katedra za hirurgiju, Unaizah, Kraljevina Saudijska Arabija

Uvod: U oblasti medicinskog obrazovanja, pitanja sa višestrukim izborom (MCQs) predstavljaju najčešće korišćeni metod procene. Neophodno je analizirati rezultate putem analize stavki kako bi se osigurao odgovarajući kvalitet. Ova studija je procenila kvalitet MCQ-ova koji se koriste za sumativnu procenu studenata na ispitu iz opšte hirurgije sprovedenih tokom 2023-24. godine na Medicinskom fakultetu (Unaizah), Univerziteta Qassim u Saudijskoj Arabiji.

Metode: Koristeći nekoliko uspostavljenih parametara za analizu stavki, studija je procenila pitanja sa višestrukim izborom u pogledu težine, diskriminatorne moći i kvaliteta distraktora.

Rezultati: Kvalitet pitanja je varirao. Srednje vrednosti indeksa težine, indeksa diskriminacije, efikasnosti diskriminacije i efikasnosti distraktora bile su, redom, 76.31%, 0.28, -0.7743 i 32%.

Zaključak: Analiza stavki je ključna tehnika za procenu kvaliteta MCQ-ova. Postojale su višestruke greške u MCQ-ovima korišćenim u sumativnim procenama, što otkriva prostor za dalje poboljšanje u budućim kursevima. Važno je redovno organizovati edukacije nastavnog osoblja kako bi se prenela potrebna znanja i veštine za kreiranje MCQ-ova koji su validni, pouzdani i visokog kvaliteta.

Ključne reči: Procena, Analiza stavki, Pitanja sa višestrukim izborom, Indeks težine, Indeks diskriminacije, Efikasnost distraktora, Neefikasan distraktor, Analiza distraktora.

REFERENCES

1. Holmboe ES, Sherbino J, Long DM, Swing SR, Frank JR. The role of assessment in competency-based medical education. Med Teach. 2010; 32(8): 676-82. doi: 10.3109/0142159X.2010.500704.

2. Rae MG, Abdulla MH. An investigation of preclinical medical students' preference for summative or formative assessment for physiology learning. Adv Physiol Educ. 2023; 47(3): 383-92. doi: 10.1152/advan.00013.2023.

3. Arjoon JA, Xu X, Lewis JE. Understanding the state of the art for measurement in chemistry education research: Examining the psychometric evidence. J Chem Educ. 2013; 90(5): 536-45. doi: 10.1021/ed3002013.

4. Schuwirth LW, van der Vleuten CP. Different written assessment methods: what can be said about their strengths and weaknesses? Med Educ. 2004; 38(9): 974-9. doi: 10.1111/j.1365-2929.2004.01916. x.

5. Kar SS, Lakshminarayanan S, Mahalakshmy T. Basic principles of constructing multiple choice questions. IJCFM. 2015; 1(2): 65-9. doi: 10.4103/2395-2113.251640.

6. Towns M.H. Guide to developing high-quality, reliable, and valid multiple-choice assessments. J Chem Educ. 2014; 91(9): 1426-31. doi: 10.1021/ed500076x.

7. Tarrant M, Ware J, Mohammed AM. An assessment of functioning and non-functioning distractors in multiple-choice questions: A descriptive analysis. BMC Med Educ. 2009; 9: 40. doi: 10.1186/1472-6920-9-40.

8. Hingorjo MR, Jaleel F. Analysis of one-best MCQs: the difficulty index, discrimination index and distractor efficiency. J Pak Med Assoc. 2012; 62(2): 142-7.

9. Odukoya JA, Adekeye O, Igbinoba AO, Afolabi A. Item analysis of university-wide multiple choice objective examinations: the experience of a Nigerian private university. Qual Quant. 2018; 52(3): 983-97. doi: 10.1007/s11135-017-0499-2.

10. Frey BB, Petersen S, Edwards L M, Pedrotti J T, Peyton V. Item-writing rules: Collective wisdom. Teaching and Teacher Education, 2005; 21: 357-64. doi: 10.1016/j.tate.2005.01.008.

11. Mc Coubrie P. Improving the fairness of multiple-choice questions: a literature review. Med Teach. 2004; 26(8): 709-12. doi: 10.1080/01421590400013495.

12. 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): 133-43. doi: 10.1007/s10459-004-4019-5.

13. 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. Sci Rep. 2015; 5: 9556. doi: 10.1038/srep09556.

14. Lai H, Gierl MJ, Touchie C, Pugh D, Boulais AP, De Champlain A. Using Automatic item generation to improve the quality of MCQ distractors. Teach Learn Med. 2016; 28(2): 166-73. doi: 10.1080/10401334.2016.1146608.

15. Chauhan P, Chauhan GR, Chauhan BR, Vaza JV, Rathod SP. Relationship between difficulty index and distracter effectiveness in single best-answer stem type multiple choice questions. Int J Anatomy Res. 2015; 3(4): 1607-10. doi: 10.16965/ijar.2015.299.

16. Patel R M. Use of Item analysis to improve quality of Multiple-Choice Questions in II MBBS. J EducTechnol Health Sci. 2017; 4(1): 22-9.

17. Mahjabeen W, Alam S, Hassan U, Zafar T, Butt R, Konain S, et al. Difficulty index, discrimination index and distractor efficiency in multiple choice questions. Ann Pak Inst Med Sci. 2017; 13(4): 310-5. doi: 10.48036/apims.v13i4.9.

18. Rao C, Prasad HK, Sajitha K, Permi H and Shetty J. Item analysis of Multiple- Choice Questions: Assessing an assessment tool in medical students. Int J Educational Psychol Res. 2017; 2(4): 201-4. doi: 10.4103/2395-2296.189670.

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 J Community Med. 2014; 39(1): 17-20. doi: 10.4103/0970-0218.126347.

20. Lama C P, Kharbuja R, Karki D, Dhungel S. Study on item analysis of Multiple-Choice Questions amongst the undergraduate dental students. Nepal Med Coll J. 2023; 25 (4): 301-5. doi: 10.3126/nmcj.v25i4.60876.

21. Bhat SK, Prasad KH. Item analysis and optimizing multiple-choice questions for a viable question bank in ophthalmology: A cross-sectional study. Indian J Ophthalmol. 2021; 69(2): 343-6. doi: 10.4103/ijo.IJO_1610_20.

22. Baste VS. Item analysis of MCQs in physiology and its correlation with faculty's perception of difficulty level of MCQs. Natl J Physiol Pharm Pharmacol. 2023; 13 (7): 1444-8. doi: 10.5455/njppp.2023.13.12601202220122022.

23. AlFaris E, Naeem N, Irfan F, Qureshi R, Saad H, Al Sadhan R, et al. A one-day dental faculty workshop in writing Multiple-Choice Questions: An impact evaluation. J Dent Educ. 2015; 79(11): 1305-13.

24. Ebrahimi S, Kojuri J. Assessing the impact of faculty development fellowship in Shiraz University of Medical Sciences. Arch Iran Med. 2012; 15(2): 79-81.

25. Elliot DL, Skeff KM, Stratos GA. How do you get to the improvement of teaching? A longitudinal faculty development program for medical educators. Teach Learn Med. 1999; 11: 52-7. doi: 10.1207/S15328015TLM1101_12.

Correspondence to/Autor za korespondenciju
Lamees Al Sulaim
Assistant Professor, Department of Surgery, College of Medicine, Qassim University,
Kingdom of Saudi Arabia
email: Dr.lameesz@gmail.com; L.alsulaim@qu.edu.sa
ORCID:0000-0002-5785-8468
Research ID:GXH-4210-2022
SCOPUS ID: 57900682200
Present address: Qassim, Qassim University Buraidah 52571

***How to cite this article**:* Salati AS, AlSulaim SL. Item analysis of Multiple-Choice Questions in an undergraduate surgery course – an assessment of an assessment tool. Sanamed. 2024; 19(2): 163-171. Doi: 10.5937/sanamed0-50691.